CUH-HO 5938-15-KP25

Calcutta University Library
ALIPORE CAMPUS
22-08-05

THE JOURNAL OF PHILOSOPHY

18

VOLUME CII, NUMBER 1
JANUARY 2005

page
5 *Supervenience and Object-dependent Properties* Thomas Hofweber
33 *Functionalism at Forty: A Critical Retrospective* Paul M. Churchland

51 MEMO TO AUTHORS
52 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

KP 2563

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc.

ISSN 0022–362X

Calcutta University Library
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

KP 2563

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J. E. Woodbridge and J. McKeen Cattell. In 1906, Wendell T. Bush became associated with the JOURNAL as co-editor. The title was changed in 1921, and the JOURNAL was incorporated in the State of New York under its present name in 1923. Since its founding, it has been published from the Department of Philosophy at Columbia University. It is typeset and printed by Capital City Press.

THE JOURNAL OF
PHILOSOPHY

VOLUME CII, NUMBER 1
JANUARY–DECEMBER 2005

Edited by
Bernard Berofsky, Akeel Bilgrami
John Collins, Arthur C. Danto
Kent Greenawalt, Patricia Kitcher
Philip Kitcher, Isaac Levi, Wolfgang Mann
Mary Mothersill, Philip Pettit
Carol Rovane, Achille C. Varzi

Consulting Editors
David Albert, James T. Higginbotham
Charles D. Parsons, Wilfried Sieg

Managing Editor
John Smylie

Published by The Journal of Philosophy, Inc.

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers. The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use. Accordingly, the JOURNAL authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $.05 per page/per copy is paid directly to us. *Each copy must include a notice of copyright.* We shall turn over to the authors any substantial payments (over $30.00, an administrative fee) obtained from all photocopying. Inquiries about photocopying may be sent to the Managing Editor.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and other mailing offices. Please send address changes to THE JOURNAL OF PHILOSOPHY, 709 Philosophy Hall, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. Fax: (212) 932-3721

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NO. 1, JANUARY 2005

SUPERVENIENCE AND OBJECT-DEPENDENT PROPERTIES*

It is hardly ever doubted that everything that exists in space and time is made up from stuff which is ultimately physical. Everything there is in space and time is ultimately physical stuff, either complex or simple. But whether all the properties of the physical stuff are ultimately physical properties, either complex or simple, is much more controversial. Most philosophers, I think it is fair to say, suspect that the answer is no. But giving up that all properties of physical things are physical properties is not to give up a physicalist world view. The standard way to reconcile nonphysical properties of physical objects with a physicalistic world view is to claim that the physical properties determine all the properties. Thus the physical properties of a physical object determine what nonphysical properties that physical object has. Or to put it differently, the nonphysical properties of physical objects supervene on their physical properties. This might not be sufficient to defend a physicalistic world view since by itself it leaves many questions unanswered: How do the physical properties determine the nonphysical ones? Why do particular nonphysical properties supervene on particular physical properties? And so on. However, the supervenience of the nonphysical properties of physical objects on their physical properties is generally believed to be a necessary condition for a satisfactory physicalistic world view, and an acceptable version of physicalism.

In this paper I will argue that the supervenience of the nonphysical

---

* This paper profited from conversations with Kent Bach, Karen Bennett, David Chalmers, Anthony Everett, Avrom Faderman, Terry Horgan, Jim Joyce, Marc Lange, John Perry, David Pitt, Ian Proops, Stephen Schiffer, David Velleman, and Jessica Wilson. Thanks for all your help! Earlier versions were presented at the Stanford-Berkeley conference in 1996 and at the Pacific Division American Philosophical Association meeting in Albuquerque in 2000. Thanks to Josh Dever and Rod Bertolet for their thoughtful comments on these occasions.

0022-362X/05/0201/5–32

on the physical is incompatible with a certain metaphysical view about properties, the view that there are object-dependent properties. This latter view, in turn, is implied by a widely held view in the philosophy of language, the thesis of direct reference. Thus overall I will argue that direct reference and supervenience of the nonphysical on the physical are incompatible.

It might seem surprising how one could think that there is a conflict between direct reference and the supervenience of the nonphysical on the physical. After all, supervenience is a view in metaphysics and direct reference is a view in the philosophy of language. But such a conflict does arise because of a consequence about properties that direct reference has. Direct reference guarantees, as we will see, that some physical objects have object-dependent properties. And some of these properties have a special general feature, which I will capture with a notion to be defined below, that of a local uniqueness property. I will argue that local uniqueness properties cannot supervene on physical properties. To establish that there are local uniqueness properties and that some physical objects have them is the only role that direct reference and object-dependent properties play in this paper. The whole argument to follow could be spelled out by not mentioning direct reference at all, but instead by making the assumption explicit that there are local uniqueness properties. However, this assumption would certainly seem dubious, implausible, and farfetched, and therefore the whole argument will appear to be of little interest. By making clear that this assumption follows from the widely held thesis of direct reference, this can be avoided, and the overall conclusion will seem of much greater relevance. Thus my argument is an argument that direct reference and the supervenience of the nonphysical on the physical cannot both be true. This should be of interest since many philosophers, myself included, would like to believe in both. But if the argument to come is correct then there is a serious tension between these two views.

To establish my main conclusion that direct reference and supervenience are incompatible I will first have to say a little bit about direct reference, and then a little bit more about supervenience. After that I will establish, using direct reference as an assumption, that certain properties are local uniqueness properties. Finally, I will argue that if supervenience of the nonphysical on the physical is true then no property of a physical object can be a local uniqueness property. This will occupy most of this paper, and it is my main conclusion. The argument that establishes this conclusion can also be used to shed some light on an old debate about the identity of indiscernibles and its relationship to properties which are a *thisness* or *haecceitas*. I will try to illuminate this debate in section v, after the main argument has been presented. Finally I will discuss some attempts to get around the main argument, and what lessons should be learned from all this.

I. DIRECT REFERENCE AND OBJECT-DEPENDENT PROPERTIES

Direct reference is a thesis in the philosophy of language. It is concerned with the semantic function of certain singular terms, in particular proper names. The thesis has many subtle features and notorious problems, but they will not take center stage in this discussion.[1] The most important thing here is the relationship between direct reference and object-dependent properties and propositions. To make this explicit I will characterize direct reference as follows: direct reference is true if some terms in our language are directly referential. A term is directly referential if the contribution it makes to the content of an utterance in which it occurs is merely the object referred to, as opposed to a mode of presentation, some description, or some other characterization of that object. In particular, the proposition expressed by an utterance containing a directly referential term will be an object-dependent proposition, it depends on the object that the directly referring term contributes. A proposition can depend on an object in at least two different ways. First, a proposition can depend on an object for its identity. That is, if Jones and Fred are one and the same then the proposition expressed by

(1) Jones is hungry.

and the proposition expressed by

(2) Fred is hungry.

are also one and the same. If the objects referred to are identical, and if there is no other relevant difference between utterances (like different disambiguation, and so on) then the propositions are identical. It only matters what object is referred to, not how it is referred to.

Second, a proposition can depend on an object for its existence. If a proposition depends on an object for its existence then the proposition exists only if the objects exist. So, if Jones does not exist then the object-dependent proposition that Jones is hungry does not exist either.

Usually direct reference is taken to imply object dependence for the propositions in both of these senses of object dependence. Since the directly referring term contributed to the proposition expressed only the object referred to, it only matters what object is referred to, not how it is referred to. Thus we get dependence for identity. And since the object referred to is all that is contributed, if there is no object that is referred to, because what one attempts to refer to does not exist, then there is no complete proposition. In particular, the (complete) proposition does not exist if the object that one attempts to refer to does not exist. We do not have to debate here what things propositions are, but one way of looking at propositions nicely illustrates the point of the dependence of a proposition on individual objects both for identity and for existence. A Russellian view of propositions holds that they are structured entities, ordered sets that can contain regular objects as members. According to this view the proposition expressed by, say, (1) is the ordered pair consisting of Jones himself and the property of being hungry. If Jones does not exist then this set does not exist, and if Jones is Fred then the ordered pair consisting of Jones and a property is the same as the ordered pair consisting of Fred and that property. A Russellian theory of propositions illustrates object dependence in both senses, but we do not have to subscribe to it here as the correct model or account of object-dependent propositions.

---

[1] For more details see François Récanati, *Direct Reference: From Language to Thought* (Cambridge: Blackwell, 1993).

Propositions are what is expressed by utterances of full sentences. Properties, on the other hand, are what is expressed by utterances of predicates. And what was said about object-dependent propositions applies, mutatis mutandis, to object-dependent properties. Thus utterances of predicates with a directly referring term in them express object-dependent properties, and a property can depend on an object for its identity and for its existence. So, if direct reference is true then an utterance of

(3) is Jones's brother

expresses an object-dependent property, the same property as is expressed by

(4) is Fred's brother

Since supervenience is usually formulated by taking recourse to properties rather than propositions I will mainly talk about properties in the following. However, one could just as well formulate supervenience in terms of propositions, or facts, or the like. The same arguments, mutatis mutandis, will go through.

Direct reference strictly understood is merely a thesis about what properties or propositions are expressed by our utterances of predicates or sentences. Strictly understood this is compatible with a global semantic error thesis, the thesis that no utterances of sentences with directly referring terms in them are true, and that no predicates with such terms in them truly apply to anything.[2] Strictly speaking these are separate issues, but everyone who believes in direct reference believes that some utterances containing directly referring terms express true propositions. Which ones these are is, of course, a further question. So, I take the thesis of direct reference to contain a denial of global semantic error. Thus if direct reference is true then some sentences with directly referring terms in them can be truly uttered, and some predicates with directly referring terms in them truly apply to some objects. Simply put, some object-dependent propositions are true, and some objects have object-dependent properties.

This explicates the notions of direct reference and object-dependent property, as they will be used in the rest of this paper. The further notion to be clarified in my argument that direct reference and supervenience cannot both be true is the notion of supervenience.

II. SUPERVENIENCE

II.1. The Idea of Supervenience. More has to be said about the supervenience, and in particular the supervenience of the nonphysical on the physical. To understand it better we have to look at how it is motivated. This is often done with the following (undoubtedly physically naive, but nonetheless very powerful) consideration: the world contains exclusively objects that are built up from smallest parts or particles. These smallest particles have certain basic properties and relations to each other. In a sense, that is all there is. Once it is fixed what smallest particles there are and what their basic properties and relations are then, in a sense, everything is fixed. Not every property of smallest particles is basic, though. That a certain particle is within a mile of a chair is one of its properties, but it is not a basic one. The basic properties and relations are only the ones that are both relevant for the behavior of the particle over time, and also jointly sufficient for that particle to have that behavior (in fixed circumstances). Physics is supposed to tell us what these basic properties and relations are and how they determine the behavior of objects that consist exclusively of these smallest particles. These basic properties and relations are thus physical properties and relations. And this gives rise to the thought that the physical fixes the rest. In a world that consists only of these smallest particles and complexes of them, everything is fixed once it is fixed what kinds of particles there are and what their basic properties and relations are. In such a world only these can really make a difference. To use a helpful analogy, after God created the physical objects and determined their basic properties and relations to each other, there was nothing else to do. So, once the physical is fixed everything is fixed. The physical determines the rest. Or in other words, the nonphysical supervenes on the physical.

---

[2] One might think that negation and negative properties alone guarantee that some propositions are true, but this will require further argument. I would like to sideline this issue here and rather make the denial of global semantic error explicit.

This is a motivation for supervenience, but it is not an account of supervenience yet. Supervenience of the nonphysical on the physical is the claim that the physical determines the nonphysical. How "determination" should be understood here still has to be accounted for. To spell this out more precisely is the job of a supervenience principle. Such a principle is supposed to provide at least a necessary condition for determination, and would ideally also give us a sufficient condition. A variety of supervenience principles have been discussed in the literature, and their relationship has been closely investigated. I will only repeat as much of this discussion as is necessary for the main point of the present paper.

One trivially sufficient condition for determination is reduction. There are different notions of reduction, but it will not be necessary to look at the details of different conceptions of reduction here. In the easiest case, nonphysical properties can be reduced to physical properties in the sense that nonphysical properties of physical objects (smallest particles or complexes thereof) are just complex physical properties. If the nonphysical can be reduced to the physical in this sense then trivially the physical determines the nonphysical. But one of the main motivations for the close study of supervenience is the possibility of a version of nonreductive physicalism. Reduction surely is sufficient for determination, but it might not be necessary. The key to defending a nonreductive form of physicalism is to spell out and defend a supervenience principle that is consistent with the falsity of reductionism.

Several such principles have been proposed, the most prominent of which are *global supervenience* (GS) and *strong supervenience* (SS):

(GS) If two possible worlds are physically indistinguishable then they are nonphysically indistinguishable.

(SS) Necessarily, for every nonphysical property *N* that a physical object *O* has, there is a physical property *P* that *O* has, such that: necessarily, if something has *P* then it has *N*.

It was originally argued by Jaegwon Kim that they are equivalent, but that was a mistake. If we are very generous in what conditions properties are closed under, then (SS) implies (GS), but not the other way round.[3] However, (SS) is equivalent to[4]

> (SS_2) For any two possible objects, if they have the same physical properties then they have the same nonphysical properties.

With most of the supervenience principles formulated in terms of properties, the formulation of them implicitly relies on certain conceptions of properties, at least on what properties there are. For example, believers in *property elitism*, the view that there are only view and special properties, and that properties are not closed under many operations that predicates are closed under, like conjunction, disjunction, and negation, will have to formulate supervenience differently, either by not taking recourse to properties at all, or by taking recourse to sets of properties.[5]

These supervenience principles have faced several objections challenging that they will be able to play a core role in a defense of a form of nonreductive physicalism. First, it has been argued that these principles are too strong since they imply reductionism. For example, Kim has argued that (SS) implies reductionism.[6] This argument has been widely criticized, in particular for its relying on very strong closure conditions on properties. This discussion has created a large literature, but since my main argument is independent of this debate I will not pursue it further here.

Second, (SS) has been accused of being too strong, unless the notion of necessity involved in it is understood as being less strong than metaphysical necessity. After all, if the laws of physics are contingent, or if what the physical properties are differs from one world to the next, then we cannot expect (SS) to hold with "necessity" being understood as metaphysical necessity, but we have to understand it as nomological, or physical necessity. This issue is also important and interesting, but it is again of no importance for this discussion. We can take any notion of necessity for what is to come, in fact, we could even leave the first "necessarily" out in (SS) in the discussion below.

---

[3] See Kim, "Concepts of Supervenience," and "'Strong' and 'Global' Supervenience Revisited," in *Supervenience and Mind* (New York: Cambridge, 1993), pp. 53–78 and 79–91; see also Bradford Petrie, "Global Supervenience and Reduction," *Philosophy and Phenomenological Research*, XLVIII (1987): 119–30.

[4] See Kim, "'Strong' and 'Global' Supervenience Revisited." This, again, assumes generous closure conditions on properties.

[5] A well-known sparse theory of properties is David Armstrong's, for example in *Universals: An Opinionated Introduction* (Boulder, CO: Westview, 1989).

[6] See Kim, "Concepts of Supervenience."

In the following any notion of necessity will do, and so I will not pursue this further.

Third, it has been argued that these principles are too weak, since even though their truth is necessary for the determination of the nonphysical by the physical, they are not sufficient for such a determination.[7] This is usually supported with the following consideration. The above principles only claim that physically indistinguishable objects or worlds are nonphysically indistinguishable. It is consistent with this that physically very similar but distinguishable objects or worlds have completely different nonphysical properties. For example, it is consistent with these principles that in a world just like ours, except that there is one more pebble on a far away planet, the nonphysical properties of physical objects are completely different than they are in our world. The above principles only require that physically indistinguishable objects or worlds are indistinguishable with respect to their nonphysical properties. But all physical objects in the extra-pebble-world are physically distinguishable from the corresponding objects in our world, and the world as a whole is physically distinguishable from our world. They, for example, have the property of being one million miles from a planet with 1001 pebbles, rather than the property of being one million miles from a planet with 1000 pebbles. It is consistent with (GS), (SS), and (SS_2), that the nonphysical properties of the corresponding objects are completely different.[8] It has thus been claimed that such principles need to be strengthened in some way or other to be closer to capturing determination. The above principles might be necessary for determination of the nonphysical by the physical, but they are not even close to being sufficient. In fact, I have heard them being accused of being basically vacuous.

Several attempts have been made in the literature to strengthen supervenience so that it can more plausibly be taken to capture determination. One was proposed by Kim.[9] He proposed not only to connect physical indistinguishability with nonphysical indistinguishability, but also physical similarity with nonphysical similarity. Whether this is satisfactory is quite controversial, however, since it is not at all clear that an account of determination has to rule out that small physical changes cannot lead to large nonphysical changes.

---

[7] See, for example, Terence Horgan, "Supervenience and Microphysics," *Pacific Philosophical Quarterly*, LXIII (1982): 29–43; Kim, "Concepts of Supervenience," and "'Strong' and 'Global' Supervenience Revisited."

[8] See Kim, "The Myth of Non-reductive Materialism," in *Supervenience and Mind*, pp. 265–84, pp. 277ff. The argument has been criticized by R. Cranston Paull and Theodore Sider in "In Defense of Global Supervenience," *Philosophy and Phenomenological Research*, XLII (1992): 833–54; and Robert Stalnaker in "Varieties of Supervenience," *Philosophical Perspectives*, X (1996): 221–41.

[9] In "'Strong' and 'Global' Supervenience Revisited."

Another attempt to get closer to determination is by taking recourse to locality. Intuitively, not everything in the universe is relevant for, say, my chair's being a chair. What matters for that is only what is going on inside a small region of the universe. This could be made more precise in a supervenience principle that takes recourse to what I will call an *unconditional locality requirement*. This is a requirement that only a limited region of the universe alone has to determine some physical object's having a nonphysical property. To turn this into a supervenience principle one could modify global supervenience and claim that if two regions (within a world, or across worlds) are physically indistinguishable then they have to be nonphysically indistinguishable. Global supervenience would be a limit of this, where the regions are complete worlds.[10]

But this gives rise to at least two problems. First, what is it for a region to be indistinguishable from another? Can we not take recourse to relational properties of these regions, and thus do we not collapse such supervenience principles into global supervenience? Second, why should we think that such a principle is true, assuming it can be formulated in such a way that it does not collapse into global supervenience? After all, why should it be guaranteed that physically indistinguishable regions are indistinguishable, in an interesting sense of the notion, in all nonphysical respects? Are not the heads of Oscar and his twin physically indistinguishable (as always, we ignore the fact that brains contain water), but one believes that water is wet, while the other believes XYZ is wet? Locality understood as unconditional locality thus will not help us.

II.2. Conditional Locality and Supervenience. To require for an account of determination that regional physical indistinguishability brings with it nonphysical indistinguishability is too strong. But a weaker form of local dependence can be required for determination. Intuitively we can motivate it as follows. Take my chair's having the property of being a chair. And take some region of the universe, say, the solar system we are in. It is not a requirement of the notion of determination that my chair's being a chair is determined by the physical features of the solar system alone. It is conceivable that the whole physical universe is needed to determine that my chair has the property of being a chair. But if my chair's being a chair only depends on what is going on inside the solar system then the whole physical universe cannot be required for the determination of my chair's being a chair. If my chair's being a chair only depends on what is going on inside the solar system then the physical properties of my chair that are such that my chair's having them also only depends on what is going on inside the solar system alone have to be sufficient to determine that my chair has the property of being a chair. To put it more simply, if my chair's being a chair only depends on what is going on inside the solar system then only its physical properties that themselves only depend on what is going on inside the solar system have to determine that it is a chair. (I will introduce some terminology momentarily that will allow me to put it much more simply.) Thus we do not require there to be unconditional local dependence. It is not claimed that my chair's being a chair only depends on what is going on inside the solar system. Rather, the claim is that *if* my chair's being a chair only depends on what is going on inside the solar system, then the physical properties that likewise only depend on what is going on inside the solar system have to determine it.

---

[10] Locality considerations as strengthenings of supervenience principles were proposed by Horgan in "Supervenience and Microphysics," and "Form Supervenience to Superdupervenience: Meeting the Demands of a Material World," *Mind*, CII (1993): 555–86. His "regional supervenience" is more properly understood as a conditional locality principle, as explained below.

I will try to formulate this idea more precisely in this section and will refine it further in later sections. Harmless as it sounds, it will be central in the main argument of this paper.

To have a supervenience principle that captures this conditional local dependence I will introduce the notion of a local property. With it we can then refine (SS). For simplicity, let us take some region that is large enough, but still less than all of the universe, like our solar system. We can say that a property is local (relative to the solar system) just in case something inside the solar system has it, and that thing would still have it as long as everything inside the solar system remains the same. More precisely:

(5) A property, physical or nonphysical, is *local* iff a physical object inside the solar system has that property and that object would still have this property even if the universe outside the solar system were different, but the same inside the solar system (over time).[11]

Now, if the physical determines the nonphysical, and if there is a local nonphysical property that a physical object inside the solar system has, then the *local* physical properties of that object have to determine that nonphysical property. Since, after all, how could the physical determine the nonphysical, a physical object's having a certain nonphysical property only depend on what's going on inside the solar system, but the physical properties that determine the physical object's having that nonphysical property not be local?

---

[11] This explication of being a local property will be expanded on below.

We can simply modify (SS) to accommodate this idea. The resulting principle could be called *Local-Local Supervenience*:[12]

> (LLS) Necessarily, for every local nonphysical property *N* that a physical object *O* (inside the solar system) has, there is a local physical property *P* that *O* has, such that: necessarily, if something has *P* then it has *N*.

(LLS) is on the right track when it comes to getting closer to capturing determination with a supervenience principle. Of course, (LLS) is not the strongest principle that could be motivated. It is restricted to one region, the solar system, which could easily be strengthened in the same spirit.[13] But (LLS) is better than (SS) and (GS). It makes explicit something that cannot be denied if one believes in determination of the nonphysical by the physical.

If the nonphysical is determined by or supervenes on the physical then (LLS) will have to be true. This is not to endorse unconditional locality, just conditional locality. Only local properties have to supervene on local properties, not necessarily all properties.

But here is the rub: (LLS) cannot be true if direct reference is true. If direct reference is true then there are object-dependent properties and physical objects have them. But if physical objects have object-dependent properties then (LLS) fails. Thus if direct reference is true then supervenience cannot be true.

---

[12] Horgan's regional supervenience is a related supervenience principle. See his "Supervenience and Microphysics," and "Form Supervenience to Superdupervenience: Meeting the Demands of a Material World." It should also be noted that (LLS) only applies to local properties, and does not imply (SS) in general, and thus is not a substitute for (SS), only an augmentation. Since local properties are the primary concern here I will work with this formulation. Of course, in the general setting (SS) is still assumed for nonlocal properties.

[13] An obvious strengthening that is in the spirit of how (LLS) was motivated is as follows: first define being a local property with respect to a space-time region *R*:

> (6) A property is *R*-local just in case something in *R* has it and that thing would still have it even if the universe outside of *R* were different, but the same inside of *R* (over the duration of *R*).

Then define a *stronger* version of local-local supervenience:

> (SLLS) Necessarily, for every region *R* and for every nonphysical *R*-local property *N* that a physical object *O* inside *R* has, there is a *R*-local physical property *P* that *O* has, such that: necessarily, if something has *P* then it has *N*.

III. THE ARGUMENT

First we have to have a look at which properties are local. We can do so by considering a certain conditional, the one that defines being a local property. This makes it easy to see that the property of being a chair is local. My chair has it and it would still have it as long as the world inside the solar system would be the same, even if it would be different outside of it. The property of observing a certain faraway star is not local, though, even if something inside the solar system has it. If we leave everything inside the solar system untouched but take away that faraway star then whatever had the property of observing that star before would no longer have it.

Now consider an object-dependent property, like the property of being Clinton. This property, as an object-dependent property, is simply the property of being identical to a particular object. Since Clinton is in the solar system, and since he has the property of being identical to Clinton, we have to ask ourselves: Would this object still have this property if we changed the world outside of the solar system, as long as we leave everything the same inside of it? Since the property of being Clinton is merely the property of being identical with a certain object, changing around which objects there are outside of the solar system and which properties they have has no effect on which objects there are inside the solar system. Clinton would still be there and he would still have the property of being identical to Clinton. This might seem trivial, and it is. But it is only trivial if we assume that the property of being Clinton is an object-dependent property. If it were the property of being the *F*, for some purely general *F*, as certain descriptive theories would hold, then this is not trivial any more, and is in fact false. If the property of being Clinton is the property of being the *F* then we could change the world around in such a way that outside of the solar system we add another *F*. Thus Clinton would not be the (one and only) *F* any more. But under the assumption that the property of being Clinton is an object-dependent property it indeed follows trivially that it is a local property. Even if we add more objects outside of the solar system, or take some objects away, or change some of their properties, Clinton remains untouched, and is still identical to the very same object, Clinton. Object-dependent properties trivially have this feature, but not all properties do. The same considerations carry over to properties like being Clinton's mother and a wide range of other object-dependent properties.

But now look at Local-Local Supervenience again which was established above to be at least a necessary condition for the supervenience of the nonphysical on the physical. Since Clinton is a local physical object and being Clinton is a local physical property there thus has to be a local physical property that is such that: necessarily, if something has this physical property then it has the property of being Clinton. Let us call this physical property P_c. I will argue that there is no such physical property.

The property of being Clinton is a special property because it is not possible that more than one thing has it. Clinton might not exist in some possible worlds, but if there is something that is Clinton in some world then there is at most one such thing in that world. Let us call a property that at most one thing can have in each possible world a *uniqueness property*. Note now that the physical property P_c whose existence is claimed by (LLS) has to be a uniqueness property, too. If not then it is possible that two things in one world have P_c. But since it is necessarily so that if something has P_c then it has the property of being Clinton, in such a possible world these two things would be Clinton, which is impossible, since being Clinton is a uniqueness property. P_c thus has to be both a local property and a uniqueness property. However, no physical property is both a local property and a uniqueness property. Here is why.

There are basically two ways in which a property can be a uniqueness property:

- It could be a property that has uniqueness built in logically, so to speak, like the property of being the (one and only) *F*. These properties would be expressed by predicates containing definite descriptions, or other devices involving quantification and identity. I will call them *logical uniqueness properties*.
- Or it could be an object-dependent property, like the property of being Clinton.

However, no local physical property is either one of these.

It is easy to see that no local physical property *P* can be a logical uniqueness property. Assume that *P* is the property of being the (one and only) *F*. Since *P* is local, an object *O* that is the *F* would still be the *F* even if the universe outside the solar system were different. But, of course, if there were another *F* outside the solar system then *O* would not be the (one and only) *F* any more. Unless being *F* is itself a uniqueness property (in which case we run the same argument with being *F*) it will be possible that there is another thing that is *F* but the solar system remains the same.

In addition, no physical property is object dependent. To see this we have to look a little bit more closely at which properties are the physical properties. Not all properties of physical objects are physical properties. The whole point of the discussion of supervenience and determination is to see whether all of the properties of physical objects are determined by a special select class of properties of these objects. To see which properties are the physical ones we have to look back at what motivated the idea of physicalism, the idea that the physical determines the rest. The idea, again simply put, was that all physical objects are made up from smallest stuff, and this smallest stuff has basic properties and relations which together are sufficient to determine its behavior. In addition, the properties that count as physical are only the ones that are necessary in the determination of the behavior of the smallest stuff. If one class of properties is sufficient to determine the behavior then any class that contains it will be sufficient as well, but this larger class will contain nonphysical properties. The physical properties are thus the smallest class of properties that are sufficient to determine the behavior of the smallest stuff. The idea of physicalism is that since everything is built up from this smallest stuff these properties are sufficient to determine the rest. Thus the question whether or not some physical properties are object-dependent properties is not whether or not some physical objects have object-dependent properties, but whether or not these properties are part of the smallest class of properties that are jointly sufficient to determine the behavior of the stuff that all physical objects are made of. And it is here that the strangeness and irrelevance of object-dependent properties becomes apparent at the physical level. Even if the smallest stuff has object-dependent properties, these properties are completely irrelevant for their behavior.

Take a simple example, like a particular electron, call it Elly. Elly has a number of general properties which are not object dependent. It has a charge, mass, speed, and so on.[14] These are all properties that are relevant for its behavior. The property of having a certain charge will affect its behavior in one way, and the property of having a certain mass will affect it in another. But the property of being Elly has nothing to do with its behavior. It will behave exactly like any other electron that has the same mass, charge, and so on. Only the general properties matter for how it will interact with other particles, not which particular electron it is.

Or so things seem to be in our world. In our world only the general properties matter, not which particular particle among several otherwise indistinguishable particles we are dealing with. The physical laws of our world do not care about the individual in that sense. All that is physically relevant are the general properties of the individual, not which individual it is. It is left an open question so far whether this has to be so. Could God make the world in such a way that being Elly matters for Elly's behavior and its interaction with other matter? Maybe God could, but he would make a strange world in doing this. In this world there would be an object such that being that object, and not any other with the same general properties, matters for the interaction of this object with others. This is more than merely a world where there is a physically special object, one that has a special status among all objects there are and which exhibits special behavior. It would be a world where this object is special *because* it is that object, and not because it alone has a certain special combination of general properties. But maybe God cannot make such a world. Maybe there are some general truths about laws, or causation, or the like, that require that only the general properties matter. I want to leave this question open in this paper, since all that is required for my argument is that in a world that is physically like ours the physical properties are not object-dependent properties. Our world is such that only the purely general properties matter for how individuals interact with others at the physical level. Only general properties are physically relevant, and thus are physical properties. Thus even if physical objects have object-dependent properties, these properties are not among the physical properties of these objects.

---

[14] Again, I simplify. What if physics has no room for local properties, and what if there are no objects or particulars at the level of smallest things? As we will see below, the worse for the supervenience of the nonphysical on the physical.

The claim that object-dependent properties in fact are not physically relevant, and thus not physical properties, even if physical objects have them, is ultimately an empirical claim. It might be necessarily so, but here I only maintain that it in fact is so. Whether or not it is necessarily so has to be settled in a debate about laws of nature, or causation, or a similar debate. Whether it is in fact so has to be settled scientifically. Whether it is necessarily so can be doubted, whether it is in fact so has to my knowledge never seriously been doubted, but still, I cannot claim to have proven it. To be very strict I should claim only to have argued for a conditional claim, a conditional whose antecedent is universally believed to be true, but an empirical claim that was not established in this paper. The conditional simply is that if object-dependent properties do not matter physically then object-dependent properties do not supervene on physical properties. But since the antecedent is never doubted I can risk the more ambitious claim that object-dependent properties do not supervene on physical properties. Since direct reference guarantees that physical objects have object-dependent properties we can conclude the direct reference and supervenience are incompatible.[15]

IV. REMARKS ON THE ARGUMENT

I would like to clarify several aspects of the above argument in this section.

(1) The main argument was given for the failure of the supervenience of the nonphysical on the physical, given that there are object-dependent properties. But it clearly generalizes for any kind of supervenience where the supervening properties contain object-dependent ones, but the subvenient properties do not. The physical and nonphysical are one case of this. A similar argument can be given for versions of materialism that are more permissive than physicalism, as long as this asymmetry is maintained. Other supervenience claims are untouched by all this, like the supervenience of the moral on the nonmoral, since both classes of properties arguably contain object-dependent properties.

(2) The role of direct reference in the argument is to show that there are local uniqueness properties, and that some things have such properties. Using direct reference we can show that the property of being Clinton passes the test for being a local property, and thus for being a local uniqueness property. It would not pass this test if certain descriptivist views were right. According to such descriptivist views the property of being Clinton is just the property of being the *F* (for some *F*). This property is not a local property, as we saw above. Direct reference and object-dependent properties thus do play an essential role in the argument, however, their role is limited in the sense indicated. We could eliminate talk about them by directly assuming that the property of being Clinton or related properties are local properties. But as said at the very beginning, this assumption will seem ad hoc and suspicious. Direct reference implies it, and since it is a widely held view it was used as an explicit premise.

(3) It has been doubted that a metaphysical thesis like the supervenience of the nonphysical on the physical could possibly be in conflict with a semantic thesis like direct reference.[16] After all, how can a thesis about language have any large scale implications about the basic make-up of reality? True enough, semantics all by itself is only about languages, but semantics plus the fact that certain utterances are true can have such implications. For example, given that the

---

[15] I am indebted to an anonymous referee for suggesting substantial improvements to this part of the paper.

[16] For example, in Robert Adams, "Primitive Thisness and Primitive Identity," this JOURNAL, LXXVI, 1 (January 1979): 5–26.

predicate 'is Clinton's mother' truly applies to a certain person, certain semantic facts about that predicate can have metaphysical implications. In this case, the semantic fact of direct reference would have the implication that a certain person has an object-dependent property; and similarly for sentences or utterances. If 'Clinton is tall' is true, or is truly uttered, then semantic facts about the term 'Clinton' can have metaphysical implications, implications that the semantic facts without the extra claim about truth would not have. Thus semantics and the truth of ascriptions of predicates can have the consequence that some things have object-dependent properties. If such properties cannot supervene on physical properties (as was argued above) then further metaphysical consequences follow.

This was also preemptively addressed above by assuming a denial of a global semantic error thesis. It was assumed to be part of the theory of direct reference that some propositions expressed by sentences containing directly referring terms are true, and some predicates containing directly referring terms truly apply to some objects. This is something that no believer in direct reference has ever doubted (to my knowledge), but it has to be made explicit to address the present worry.

(4) A further angle from which the above argument could be attacked is from worries about the existence of properties. It could be argued that I assumed without justification that there is a property of being Clinton, or of being Clinton's mother, or at least that direct reference implies this. However, this is a controversial assumption, one that a nominalist about properties might deny. Thus the argument relies on controversial assumptions.

It is true that I freely made use of talk about properties in this argument, and that this might seem to be in conflict with nominalism and other controversial positions. However, this is standardly done in the discussion of supervenience. If one finds talk about properties objectionable then one will either have to give up on supervenience altogether or give it a different formulation. The latter could be done in terms of facts, propositions, predicates, or others. Whichever one will choose, the above argument will go through, mutatis mutandis. The same argument will show that there is a proposition that cannot supervene on any physical proposition, or that there is a fact that cannot supervene on any physical facts, and so on. The formulation of the above argument used talk about properties, but the argument does not depend on it. It can be restated for whatever formulation of supervenience one accepts as long as it is understood that one kind of thing (properties, facts, and the like) supervenes on the same

KP 2563

kind. I will discuss an attempt to avoid the conclusion of the argument by giving this up below.

(5) The above account of a local property should not be understood as a definition of a local property in other terms, but only as an explication of this notion. As a definition it would be unsatisfactory. A property was called local just in case something (in the solar system) has it, and that thing would still have it as long as everything in the solar system remained the same, even if there are changes outside of the solar system. But what does it mean for everything inside the solar system to remain the same? Surely if there is some change outside the solar system then everything inside the solar system will undergo some change. Everything will have different relational properties to the things outside of the solar system. One might think that how "everything inside the solar system remains the same" should be understood is that there are the same objects and that even though they might have different relational properties, they all have the same *local* properties. But now, of course, this would be circular.

This situation is analogous to the one we are in when we try to understand the notion of an intrinsic property more precisely. Intuitively this notion is clear enough, and it can be explicated by saying that an intrinsic property is one that a thing can have no matter what properties other things have. But that is not quite right since other things can have relational properties that rule out that a thing has a certain intrinsic property. More precisely, an intrinsic property is a property that something can have no matter what *intrinsic* properties other things have. But that, again, is circular as a definition. The situation about local properties is analogous to the one about intrinsic properties, but it is no worse. Being a local property can be defined in terms of being an intrinsic property. The crucial clause "the universe within the solar system remains the same" can be understood as "the universe contains the same objects within the solar system, and they have the same *intrinsic* properties." This is sufficient for our present concern.[17] If the distinction between intrinsic and extrinsic properties turns out to be indefensible then the difference between a local and a nonlocal property will likely have to go as well, together with a good chunk of contemporary metaphysics. In this case, local duplication, which is used in my main argument, will collapse into global duplication, and the assumption implicit in the main argument that a part of the world could stay the same while another part changes will turn out to be false. But the distinction between intrinsic and extrinsic properties, although hard to define precisely, seems so intuitive that it is hard to believe it is not onto something.

---

[17] For more on attempts to give definitions of "intrinsic property," see Stephen Yablo, "Intrinsicness," *Philosophical Topics*, XXVI (1998): 479–504; or Rae Langton and David Lewis, "Defining 'Intrinsic'," *Philosophy and Phenomenological Research*, LVIII (1998): 333–45.

(6) Talk about local properties might seem suspicious, in particular in the context of physicalism and physical properties. After all, has recent physics not taught us that locality cannot be assumed and that there is nonlocal physical interaction? Locality as it is used in the above discussion is not incompatible with nonlocality in physics. It only states that if there is a local nonphysical property then there has to be a local physical property which determines it. It is compatible with this that there is action at a distance or superposition, or the like. If physics, however, tells us that there are no local physical properties, then the worse for the supervenience of the nonphysical on the physical, assuming that there are object-dependent properties. Such supervenience will then fail trivially. And if physics tells us that at the smallest level there are no objects or individuals then again the worse for the supervenience of the nonphysical on the physical. Thus to talk about individual physical particles in the above argument is a concession to the physicalist, not an illegitimate assumption.

V. SYMMETRY, *THISNESS*, AND THE IDENTITY OF INDISCERNIBLES

Our discussion so far was concerned with how object-dependent properties fit into a physical world, and in particular if they can be determined by physical properties. We saw that such properties are not determined by physical properties, and in fact by any properties that are not themselves local uniqueness properties. We will look at what consequences we should draw from this in just a minute. Presently I would like to focus on how this debate relates to a different but closely related debate. This debate is one that does not directly involve object-dependent properties, but properties like the property of being Clinton, whether or not they are object dependent. A property that is the property of being identical to a particular individual, like the property of being Clinton, is called a *thisness* or *haecceitas*. It is widely discussed how such properties relate to purely general or qualitative properties. Some philosophers, most notably Leibniz, believed that every *thisness* can be analyzed in terms of purely qualitative properties. Others have argued that such an analysis is impossible.[18] The main argument for this impossibility is one using complete symmetry. In a completely symmetric world mirror images will share their purely general properties, but not their *thisnesses*. However, this argument is more problematic than it might at first seem. Those who believe that an analysis of a *thisness* in terms of purely qualitative properties is possible will also believe in the identity of indiscernibles, the principle that if *x* and *y* have the same purely qualitative properties then they are identical. After all, if an analysis of a thisness is possible in terms of purely qualitative properties then having the same purely qualitative properties is sufficient for identity. Complete symmetry is impossible for believers in the identity of indiscernibles. The question then becomes whether or not it is legitimate to use complete symmetry to give an *argument* against the identity of indiscernibles and the analysis of *thisness* in terms of purely qualitative properties without begging the question against the believer in these principles.[19] Ian Hacking, for example, has argued that every nonquestion-begging description of a world that is an alleged counterexample to the identity of indiscernibles can be reinterpreted by a believer in the principle using non-Euclidean geometry in such a way that it is consistent with this principle.[20] Whether or not complete symmetry can be used to argue against the identity of indiscernibles and the analysis of *thisnesses* in terms of purely qualitative properties remains controversial.

---

[18] See Adams, "Primitive Thisness and Primitive Identity," for a variety of historical references on these debates.

The main argument in this paper does not assume a premise from this debate. It never relied on complete symmetry nor the falsity of the identity of indiscernibles. The argument goes through whether these principles are true or false. All that was necessary for the above argument to work is that local duplication is possible. This can be so even if the two local duplicates are distinguishable by their nonlocal purely qualitative relational properties. Thus local duplication is compatible with the identity of indiscernibles and with the impossibility of complete symmetry. In addition, the argument given above does not rely on believing in complete symmetry, and thus is stronger than arguments that use complete symmetry to argue for the failure of supervenience of properties about particular individuals.

My main argument does have implications for this debate, however. It shows that an analysis of a *thisness* in terms of purely qualitative properties is impossible, without relying on complete symmetry, but given two other assumptions. We should look at this last point more closely.

The question whether or not a *thisness* can be analyzed in terms of purely qualitative properties is a real question whether or not one believes that these properties are object-dependent properties. If one believes that the property of being Clinton is identical to the property of being the *F*, for some *F* which is not object dependent, then the question remains if the property of being the *F* can be analyzed in purely qualitative terms. On the other hand, if one believes that the property of being Clinton is object dependent[21] then again there remains the issue if it nonetheless can be analyzed in terms of purely qualitative properties. For us to understand this better I will have to say more about what it is to analyze a property in terms of purely qualitative properties. One obvious way to explicate this is to claim that there has to be property identity. Being Clinton is analyzable in terms of purely qualitative properties if it is identical with a purely qualitative property. A weaker requirement is determination. The properties can be different, but one has to determine the other.

---

[19] See Max Black, "The Identity of Indiscernibles," *Mind*, LXI (1952): 153–64; Ian Hacking, "The Identity of Indiscernibles," this JOURNAL, LXXII, 9 (May 8, 1975): 249–56; and Adams, "Primitive Thisness and Primitive Identity."

[20] See Hacking, "The Identity of Indiscernibles," for the argument.

The above argument can be used to argue that a *thisness* is not analyzable in terms of purely qualitative properties if we make the following two assumptions:

(1) An analysis of property *P* as property *Q* at least requires that *Q* determines *P*
(2) A *thisness* is a local property

Assumption 1 is a rather weak requirement for analysis. Assumption 2 is guaranteed, as was argued above, if a *thisness* is an object-dependent property. But in effect this assumption is weaker than to assume that a *thisness* is object dependent since there might be properties that are not object dependent, but still local. All that was required for the above main argument to get off the ground was that there are local uniqueness properties. The thesis of direct reference was taken to imply that there are object-dependent properties of physical objects, and it was argued that object-dependent properties can be local uniqueness properties. Only the last of these was then used in the argument. If a *thisness* is a local property then our argument gets going, whether or not it is object dependent. I take it to be trivial that a *thisness* is a uniqueness property.

So, given these two assumptions we can argue that a *thisness* cannot be analyzed in terms of purely qualitative properties without relying on the falsity of the identity of indiscernibles or on complete symmetry. Using the main argument above this is now quite straightforward: by assumption 1, if an analysis of a *thisness* in terms of purely qualitative properties is possible then these purely qualitative properties have to at least determine the *thisness*. By assumption 2 a *thisness* is a local property, and thus a local uniqueness property. But by the main argument above, without relying on complete symmetry or on the falsity of the identity of indiscernibles, we have seen that no purely qualitative properties can determine a local uniqueness property. Thus an analysis of a *thisness* in terms of purely qualitative properties is impossible.[22]

---

[21] I ignore here the issue of the property being object dependent with respect to its existence, which one might take to give one a cheap route against an analysis in purely qualitative terms, since purely qualitative properties do not depend for their existence on the existence of any particular objects (according to many conceptions of properties). Dependence for identity is really what matters in this discussion.

We do not need to rely on complete symmetry in this argument. We at most need local duplication of the solar system. A local duplicate of the solar system is a collection of objects that satisfy the same purely general *local* properties that the corresponding objects satisfy in our solar system. We do not have to reject the principle of the identity of indiscernibles, but we have to reject the principle of the identity of local indiscernibles, the principle that no two objects can have the same purely qualitative local properties. This latter principle is, of course, not one that has any appeal, but it can be rejected while the principle of the identity of indiscernibles is maintained. Local duplication does not require complete symmetry. Two objects can have the same purely general local properties while differing in their purely general nonlocal properties. One of them might be a million miles from one pebble, while the other one is a million miles from two. Thus the argument given in this paper against an analysis of a *thisness* in terms of purely general properties relies on weaker assumptions than the traditional arguments using complete symmetry.

VI. A WAY OUT?

The above argument shows that supervenience and direct reference cannot both be true. It cannot be that some physical objects have object-dependent properties and that all of the properties of a physical object are determined by its physical properties. This is somewhat unsettling. The supervenience of the nonphysical on the physical is usually seen to be a minimal requirement for the truth of physicalism. And direct reference is a well-motivated, though, of course, just as physicalism nonetheless controversial, view in the philosophy of language. It will not be easy to give either one of them up. Of course, one can stick with both of them and simply reject the argument given above for some reason or other. But in these final pages we should consider whether there are other options besides these. Can we accept the argument given above, and hold onto supervenience as well as direct reference? In this section we will look more closely at one option to do this. In the concluding section I will suggest another, more radical, one.

---

[22] I am indebted to an anonymous referee for a helpful suggestion to improve this argument.

We will have to have a look again at how supervenience was understood. One of the special features of the above discussion of supervenience was taking recourse to conditional locality, which is not part of standard formulations, but I think it is a legitimate requirement for an account of determination to include a local-local relationship as (LLS). I do not want to question this now. Another thing, however, which is part of standard formulations, and which was also assumed above, is that supervenience is a relationship between two sets of the same kinds of entities. I assumed that properties supervene on properties, or facts on facts, or propositions on propositions. One strategy to get around the main argument is to give up this assumption. This strategy takes the lesson of the main argument to be that it cannot be that properties, facts, and the like, are determined by the physical members of their own kind. Thus according to this strategy we at least will have to reject what one might call *kind-kind* supervenience, the supervenience of one kind of things (properties, facts, propositions) on the same kind, as long as the subvenient kind does not contain object dependence, but the supervening kind does. But maybe this is exactly how the formulation of supervenience has to be modified to make the supervenience of the nonphysical on the physical compatible with object-dependent properties after all. Maybe the argument above relied on a too specific formulation of supervenience, and maybe this can be gotten around by giving a different formulation. I will pursue this line of thought in this section.

Object-dependent properties fail to supervene on physical properties, but in a sense they do not fail to do so by much. They almost supervene on the physical properties. Take an object-dependent property which is a local uniqueness property, like being Clinton, or being Clinton's mother. It is closely related to a nonobject-dependent relation and an object that relation is had to. Thus being Clinton's mother is closely related to the being-the-mother-of relation, and Clinton. And having the mother-relation to Clinton is closely related to having the property of being the mother of Clinton. Now, we saw that the property of being Clinton's mother, as a local uniqueness property, does not supervene on any physical property. But the relation of being-the-mother-of is not an object-dependent relation. We can grant that it supervenes on a physical relation. In addition, we can grant that Clinton is a physical object. Thus even though being Clinton's mother does not supervene on any physical property, the relation of being-the-mother-of does supervene on a physical relation, and the object the relation is had to is a physical object.

We cannot have supervenience of properties on properties, but it seems that we can have something closely related to it. An object will have the property of being Clinton's mother as long as it has the mother-of relation to Clinton. Clinton is made up from only physical stuff and is thus a physical object. In addition, there is a physical relation on which the mother-of relation supervenes on, and the object the relation is had to is a physical object, namely Clinton. So, there is a physical relation and a physical object (Clinton) such that necessarily, if something has that relation to that object then it has the property of being Clinton's mother. So, if a physical object has an object-dependent property then its having that property will be necessitated by its having a certain (nonobject-dependent) relation to another physical object (or possibly itself, as in the case of Clinton's having the property of being Clinton). If the property depends on more than one object then we need a more than binary relation and more than one physical object as relatum to do this. More generally we can formulate the following modified supervenience principle, *modified strong supervenience*:

> (MSS) Necessarily, for every nonphysical property N there is some n and an n-ary physical relation R^n, and a sequence of physical objects $\vec{o}$ of length n–1, such that, necessarily, if something has $R^n(\vec{o})$ then it has N.

This is merely a more general version of (SS). The arity of the relation can be 1, in which case the 1-ary physical relation is just a physical property, and the sequence of physical objects $\vec{o}$ will be empty. If N, however, is an object-dependent property which depends on k many objects then R will be an k+1-ary relation.

Before we will have a closer look at whether or not this is an acceptable way to make supervenience compatible with object-dependent properties, let us extend it to accommodate locality. Again, presently 'local' is understood as 'within the solar system'. A property or relation being local is understood just as we did above. An object is local if it is located within the solar system. Then we can formulate *modified local-local supervenience* as follows:

> (MLLS) Necessarily, for every local nonphysical property N there is some n and an n-ary local physical relation R^n, and a sequence

of local physical objects $\vec{o}$ of length n–1, such that, necessarily, if something has $R^n(\vec{o})$ then it has N.

This can be extended, as above in footnote 13, to cover arbitrary regions of space-time.[23]

It is now not hard to see that the modified version of local-local supervenience has no problem with local object-dependent properties. In the case of being Clinton's mother the local physical relation will be the one on which being-the-mother-of supervenes, and Clinton will be the only local physical object in the sequence of objects. In the case of the property of being Clinton, the local physical relation is identity[24] and the only physical object is, again, Clinton. That these relations are local is not hard to see with the usual test. Thus given this new formulation of supervenience we get no incompatibility between supervenience and object-dependent properties.

A second, in fact equivalent, option is to consider the notion of an objectual closure of the physical properties and relations. If we have an n-ary relation and k many objects then we can build from that (n-k)-ary object-dependent relations, namely the ones that result from putting the k many objects in certain argument places of the n-ary relation. Now, if we have a class P of properties and relations (say, the physical properties and relations) and a class O of objects (say, the physical objects) then the objectual closure of P with O are all the properties and relations that can be built from the members of P by filling argument places with members of O. In the case of the physical properties and relations, and the physical objects, the objectual closure will be the physical properties and relations plus all the object-dependent properties and relations that are such that the nonobject-dependent part consists of purely physical properties and relation, and the only objects these properties and relations depend on are physical objects. Now, we have seen above that the properties in the objectual closure of the physical properties and relations with the physical objects are not themselves physical properties. But the objectual closure of the physical properties and relations with the physical objects is sufficient as a supervenience base for the object-dependent properties (if the nonobject-dependent properties supervene on

---

[23] Here is the formulation, *stronger modified local-local supervenience*:

(SMLLS) Necessarily, for every region R and every R-local nonphysical property N there is some n and an n-ary R-local physical relation Q^n, and a sequence of R-local physical objects $\vec{o}$ of length n–1, such that, necessarily, if something has $Q^n(\vec{o})$ then it has N.

[24] As a logical relation I count this as physical.

Calcutta University ALIPORE CAMPUS

physical properties, of course). The objectual closure of the physical properties and relations will consist of object-dependent properties, and all physical objects are allowed. The above argument would be blocked.

This modification of various supervenience principles can avoid the above argument, but it does not help with the main problem. The reason is that these modified supervenience principles cannot be taken recourse to by someone who wants to spell out the determination of the nonphysical by the physical, as required for a defense of a physicalist world view. The modified supervenience principles smuggle object-dependent properties into the supervenience base, which is illegitimate in a defense of physicalism.

Physical properties are somewhat unusual. If we take the physical properties and relations, and form the objectual closure build with only physical objects then we get more than the physical properties. Physical properties are not object dependent. Even object-dependent properties that depend only on physical objects are not physical properties. In particular, the property of having a certain physical relation to a certain physical object is not a physical property. So, even though the relation is physical, and the object is physical, the property of having the relation to the object is not physical. It is important to keep this in mind. Now, we can thus see that taking recourse to the objectual closure of the physical properties and relations with physical objects is not legitimate for a physicalist supervenience base. This would classify object-dependent properties that depend on physical objects as physical properties. Thus a formulation of a modified supervenience principle taking recourse to objectual closure is no help to the physicalist, since the properties in the closure are more than just the physical properties. And the formulation of modified supervenience as in (MSS) or (MLLS) will not help either. Even though these principles do not explicitly mention object-dependent properties as part of the supervenience base, they do rely on them implicitly. The relevant clauses contain such phrases like 'there is some object...and some relation...such that having that relation to that object....' Here we quantify into the "object-position" of an object-dependent property, and thereby we are implicitly relying on object-dependent properties as the supervenience base. We are merely replacing properties that depend on particular objects with quantification into the object-position of an object-dependent property. And this is equally illegitimate for a physicalist to use as a basis for what determines the rest. The physical has to determine the rest, and since object-dependent properties are not physical they cannot be used to give an account of (physical) determination. Thus the modified supervenience principles will not help the physicalist, even though they do get around the main argument from above, since they rely on object-dependent properties in the supervenience base.

VII. CONCLUSION

Even though most of this discussion was about the physical and the nonphysical, a puzzle more generally arises about the relationship between purely qualitative properties and object-dependent properties. If we hold that objects have both purely qualitative properties as well as object-dependent properties, how do they relate to each other? The above argument showed that object-dependent properties cannot be analyzed in terms of purely qualitative properties, if analysis at least brings with it determination, a quite minimal requirement. But on the other hand, it also seems that object-dependent properties are not independent of purely qualitative properties. After all, could it really be that after God specified what kinds of things there are and what purely qualitative properties and relations these things instantiate, he still had many options left open about which objects should exist in such a world? Could God have created a world exactly like ours in all qualitative respects, but with the only difference that the thing which actually is Bush would be Clinton and the other way round? And could God thus create infinitely many worlds that are qualitatively identical and differ only in which object exist in them? It seems not.[25] But how can we maintain all that? How could the object-dependent be real and in addition to the purely qualitative, but also not be determined by it and not be independent of it?

One way out of all this, for better or worse, is as follows. The argument that led us into this situation essentially uses two kinds of modal reasoning. One of them is object centered. Here we fix the objects and consider situations where these objects have different properties. The other is world centered, or purely general. Here we merely consider what properties and relations might be instantiated and co-instantiated. We do not have to talk about any particular objects in doing this. In the main argument above both of these forms of modal reasoning were used. In the definition of being a local property, and the argument that being Clinton is a local property, we engaged in object centered modal reasoning. We imagined what would happen to particular objects if things were different. In the motivation for supervenience, on the other hand, we engaged in world centered modal reasoning. We imagined a world as a whole and wondered about what properties and relations are instantiated in such a world. Now, it might just be that these two forms of modal reasoning are not compatible with each other. This is not to say that one of them has to go. Rather it might be that even though both of these forms of reasoning are fine and important when used by themselves, when both are mixed and pushed to the limit they might just lead us into contradictions. To say this is, of course, not much more than a hint of an outline of how to deal with this situation. To spell this out we would have to look more closely at modal reasoning, what it is good for, why it comes in two kinds, and how they relate to each other. This is a big task, but if this is on the right track then it might just be that there is another way out of our dilemma.

---

[25] These worries apply to Adams's *moderate haecceitism* as well—see "Primitive Thisness and Primitive Identity."

THOMAS HOFWEBER
University of North Carolina at Chapel Hill

FUNCTIONALISM AT FORTY: A CRITICAL RETROSPECTIVE

For those of us who were undergraduates in the 1960s, functionalism in the philosophy of mind was one of the triumphs of the new analytic philosophy. It was a breath of theoretical fresh air, a framework for conceptual clarity and computational rigor, and a shining manifesto for the possibility of artificial intelligence (AI). Those who had been logical behaviorists rightly embraced it as the natural and more penetrating heir to their own deeply troubled views. Those who had been identity theorists embraced it as a more liberal but still agreeably robust form of scientific materialism. The many who hoped to account for cognition in broadly computational terms found, in functionalism, a natural philosophical home. Even the dualists who refused to embrace it had to give grudging approval for its strictly anti-reductionist stance. It had something for everyone. Small wonder that it became, and has largely remained, the dominant position in the philosophy of mind, and perhaps more importantly, in cognitive psychology and classical AI research as well.

Whether it still deserves that position—indeed, whether it ever did—is the principal subject of this essay. The legacy of functionalism, now visible to everyone after forty years of philosophical and scientific research, has not been entirely positive. But let us postpone criticism for a moment, and remind ourselves of the central claims that captured so many imaginations.

I. THE CENTRAL CLAIMS OF CLASSICAL FUNCTIONALISM

(1) What unites all cognitive creatures is not that they share the same computational mechanisms (their "hardware"). What unites them is that (plus or minus some individual defects or acquired special skills) they are all computing the same, or some part of the same, abstract <<sensory input, prior state>, <motor output, subsequent state>> *function.*[1]

(2) The central job of Cognitive Psychology is to *identify* this abstract function that we are all (more or less) computing.

---

[1] Just to remind the reader, a function is a set of input/output pairs, such that for each possible input, there is assigned a unique output. Such sets can have infinitely many input/output pairs, and the relations between the inputs and outputs can display extraordinary levels of complexity. The characterization proposed in (1) is thus in no sense demeaning to cognitive creatures. It requires only that the relevant function be computable, that is, that the proper output for any given input can be recursively generated by a finite system, such as a brain, in a finite time.

0022-362X/05/0201/33–50

(3) The central job of AI research is to create *novel physical realizations* of salient parts of, and ultimately all of, the abstract function we are all (more or less) computing.

(4) Folk Psychology—our common-sense conception of the causal structure of cognitive activity—already embodies a crude and partial representation of the function we are all (more or less) computing.

(5) The reduction of Folk Psychology (indeed, any Psychology) to the neuroscience of human brains is twice impossible, because:

(i) the relevant function is computable in a potentially infinite variety of ways, not just in the way that humans happen to do it, and
(ii) such diverse computational procedures are in any case realizable in a potential infinity of distinct physical substrates, not just in the specifically human biological substrate.

Accordingly, to reduce the categories of Folk Psychology to the idiosyncratic procedures and mechanisms of specifically *human* brain activity would be to *exclude*, from the domain of genuine cognitive agents, the endless variety of other realizations of the characteristic function (see point (1)) that we are all computing. The kind-terms of Psychology must thus be functionally rather than naturalistically or reductively defined.

(6) Empirical research into the microstructure and microactivities of human and animal brains is entirely legitimate (for certainly we do wish to know how the sought-after function is *realized* in our own idiosyncratic case). But it is a very poor research strategy for recovering the global function itself, whose structure will be more instructively revealed in the situated molar-level behavior of the entire creature.

(7) Points (5) and (6) jointly require us to respect and defend the *methodological autonomy* of Cognitive Psychology, relative to such lower-level sciences as brain anatomy, brain physiology, and biochemistry. Cognitive Psychology is picking up on its own laws at its own level of physical complexity.

Thus the familiar and collectively compelling elements of a highly influential philosophical position. Perhaps astonishingly, the position is decisively mistaken in all seven of the elements just listed. Or so, at least, I shall argue in what follows.

II. SOME UNEXPECTED LESSONS FROM NEUROBIOLOGY

The classical or "program-writing" research tradition in AI was one highly promising expression of the Functionalist view just outlined. But by the early 1980s, that research program had hit the wall with an audible thud. Despite the development of CPUs with increasingly fabulous clock speeds (even desk-top machines now top 10^9 Hz), despite ever-expanding memory capacities (even desk-top machines now boast over 10^{10} bytes), despite blistering internal signal conduction velocities (close to the speed of light), and despite the continuing a priori assurance (grounded in the Church-Turing Thesis) that a Universal Turing Machine could, in principle, compute any computable function whatever, programmed computers in fact performed very poorly relative to their biological counterparts, at least on a wide variety of typical cognitive tasks.

The problem was not that there was any well-defined class of cognitive tasks that programmed digital computers proved utterly unable to even begin to simulate. The problem was rather that equal increments of progress toward more realistic cognitive simulations proved to require the commitment of exponentially increasing resources in memory capacity, computational speed, and program complexity. Moreover, even when sufficient memory capacity was made available to cover all of the empirical contingencies that real cognition is prepared to encounter, a principled way of retrieving, from that vast store, all and only the *currently relevant* information proved entirely elusive. As the memories were made larger, the retrieval problem got worse. Accordingly, as actual cognitive performance approached the levels displayed by biological brains (and in many cases they did), the time taken for the machines to produce the desired performance expanded to ridiculous lengths. A programmed machine took minutes or hours to do what a biological brain could do in a fraction of a second.

At the time, this was deeply puzzling, because no process in the brain had a "clock frequency" higher than perhaps 100 Hz, and because typical signal conduction velocities within the brain are no greater than a human bicycle rider: perhaps 10 m/sec. In the respects at issue, this puts the biological brain at an enormous disadvantage: $\approx 10^2$ Hz versus $\approx 10^9$ Hz in the first dimension of performance, and ≈ 10 m/sec versus $\approx 10^8$ m/sec in the second. All told then, the computer should have a computational speed advantage of roughly $10^7 \times 10^7 = 10^{14}$, or fourteen orders of magnitude. And yet, as we now say, shaking our heads in amazement, the presumptive tortoise (the biological brain) easily outruns the presumptive hare (the electronic digital computer), at least on a wide variety of typical cognitive tasks.

The explanation of the human brain's impressively high performance, despite the very real handicaps mentioned, is no longer a matter of any controversy. The brains of terrestrial creatures all deploy a computational strategy quite different from that deployed in a standard serial-processing, digital-coding electronic computer. That strategy allows them to do a clever end-run around their time-related handicaps. Specifically, the biological brain is a massively *parallel* piece of computational machinery: it performs trillions of individual computational transformations—within the 10^{14} individual microscopic *synaptic connections* distributed throughout its volume—*simultaneously* and *all at once.* And it can repeat such feats of computation at least ten and perhaps a hundred times per second. The presumptive deficit of fourteen orders of magnitude scouted earlier is thus made good in one fell swoop. And the brain is left with a modest computational advantage of its own concerning the number of basic computational operations performed per second: perhaps one or two orders of magnitude over current electronic machines.

Moreover, this massively parallel, distributed processing (or PDP as it has come to be called) provides a built-in solution to classical AI's chronic problem of how to access, in real time and from the totality of one's vast memory store, all and only the informational elements that are relevant to one's current computational problem. The fact is, the acquired strengths or "weights" of the brain's 10^{14} synaptic connections collectively embody *all* of the acquired wisdom and acquired skills that the creature commands. (Learning, at least in its most basic form, *consists in* the progressive modification of those myriad connections.) But those 100 trillion synaptic connections are also the brain's basic *computational* elements. Each time a large cadre of synaptic connections effects a transformation of an incoming representation into an output representation at the receiving population of neurons, *every synapse in that entire cadre* has a hand in shaping that computational transformation, and each makes its tiny contribution simultaneously with all of the others.

Accordingly, it is not just the brain's computational behavior that is massively parallel. Its *access to memory* is also a massively parallel affair. Indeed, these are no longer distinct processes, as they are in a digital computer with a classical von Neumann architecture. In the biological brain, to engage in any computational transformation simply *is* to deploy whatever knowledge it has accumulated. Accordingly, the classical Frame Problem[2] for Artificial Intelligence simply evaporates, as does the Inductive Logician's problem of the global sensitivity (to background knowledge) of any Abductive Inference,[3] which is easily the most common form of inference that any creature ever performs.

---

[2] Daniel Dennett, "Cognitive Wheels: The Frame Problem in Artificial Intelligence," in Christopher Hookway., ed., *Minds, Machines, and Evolution* (New York: Cambridge, 1984). pp. 129–51.

[3] For a recent summary, see Jerry Fodor, *The Mind Doesn't Work That Way* (Cambridge: MIT 2000). See also, my *Inner Spaces and Outer Spaces: The New Epistemology* (in preparation), chapter 2.

These welcome developments concerning the general nature of information processing in humans and animals were humbling for the ambitions of classical AI not because those ambitions were revealed to be unachievable. On the contrary, artificial intelligence now looks more achievable than ever. Rather, these decisively illuminating developments were humbling because they were the result of empirical and theoretical research within two *lower-level* sciences, Neuroanatomy and Neurophysiology, whose contributions to Cognitive Psychology and AI were widely and officially expected to be minimal at best, and procrustean at worst. (See again points (5), (6), and (7).) But those often-derided "engineering details" turned out to be decisively relevant to understanding how a plodding biological brain could keep up with an electronic machine in the first place. And they proved equally decisive for understanding how the brain could routinely solve a vital *cognitive* problem—the real-time selective deployment of relevant information—that the programmed serial machines were quite unable to solve. Cognitive Psychology, it began to emerge, was not so "methodologically autonomous" as the Functionalists had advertised.

III. FOLK PSYCHOLOGY AS A ROUGH TEMPLATE FOR OUR COGNITIVE PROFILE: SOME PROBLEMS

More generally, the perspective on cognition that emerges from neuroanatomy and neurophysiology holds out an entirely novel conception of the brain's fundamental mode of *representation*. The proposed new unit of representation is the *pattern of activation levels* across a large population of neurons (*not* the internal sentence in some "language of thought"). And the new perspective holds out a correlatively novel conception of the brain's fundamental mode of *computation* as well. Specifically, the new unit of computation is the *transformation of one activation pattern into a second activation pattern* by forcing it through the vast matrix of synaptic connections that one neuronal population projects to another population (*not* the manipulation of sentences according to "syntactic rules"). Since our own dearly-beloved Folk Psychology *shares* in classical AI's linguaformal portrayal of human cognitive activity, the new vector-coding/vector-processing portrayal of our cognitive processes therefore casts the integrity of Folk Psychology into doubt as well, at least as an account of the *basic* structure of cognitive activity. Element (4) of the preceding functionalist manifesto is therefore severely threatened, if not outright refuted, in addition to elements (6) and (7). Its warm familiarity and yeoman social service notwithstanding, Folk Psychology appears to embody no insight whatever into the *basic* forms of representation and computation deployed by typical cognitive creatures.

This is an outcome that we should have expected in any case, since we appear to be the *only* species of cognitive creature on the planet that is capable of deploying the syntactic structures characteristic of language. If *all* cognition deploys them as the basic mode of doing business, why are the other terrestrial creatures so universally unable to learn any significant command of those linguistic structures? And if the basic modes of cognition in those other creatures are therefore almost certain to be *non*linguaformal in character, then why should we acquiesce in the delusion that human cognition—alone on the planet—*is* linguaformal in its basic character? After all, the human brain differs only marginally, in its microanatomy, from other mammalian brains; we are all closely proximate twigs on the same branch of the Earth's evolutionary tree. And the vector-coding/vector-processing story of how terrestrial brains do business is no less compelling for the human brain than it is for the brain of any other species. We have here a gathering case that Folk Psychology is a modern cousin of an old friend: Ptolemaic Astronomy. It serves the purposes of rough prediction well enough, for an important but parochial range of phenomena. But it badly misrepresents what is really going on.[4]

IV. MULTIPLE REALIZATION: ON THE ALLEGED IMPOSSIBILITY OF AN INTERTHEORETIC REDUCTION FOR ANY MOLAR-LEVEL PSYCHOLOGY

Conceivably, the preceding estimate of Folk Psychology is too harsh. Perhaps its presumptive failure to mesh with the vector-coding vector-processing story of brain activity reflects only the fact that Folk Psychology is a molar-level portrait of cognitive activity, a portrait that picks up on laws and categories at a level of description far above the details of neuroanatomy and neurophysiology, a portrait that should not be *expected* to reduce to any such lower level of scientific theory. As many will argue, that reductive demand should not be imposed on Folk Psychology—nor on any potential replacement Cognitive Psychology either (a replacement drawn, perhaps, from future molar-level research). For, it will be said, Psychology addresses law-like regularities at its own level of description. These regularities are no doubt implemented in the underlying "hardware" of the brain, but they need not be reducible to a theory of that hardware.[5] For there are endlessly many different possible material substrates that would sustain the same profile of molar-level cognitive activity.

---

[4] These skeptical themes go back a long way. See my "Eliminative Materialism and the Propositional Attitudes," this JOURNAL, LXXVIII, 2 (February 1981): 67–90. For even earlier doubts, see Paul K. Feyerabend, "Materialism and the Mind-Body Problem," *Review of Metaphysics*, XVII (1963): 49–66; and Richard Rorty, "Mind-Body Identity, Privacy, and Categories," *Review of Metaphysics*, XIX (1965): 24–54.

The claim that molar-level cognitive activities are multiply realizable is almost certainly correct. Much less certain, however, is the idea that multiple realizability counts against the possibility of an intertheoretic reduction of Folk Psychology, and against the reduction of any scientific successor Cognitive Psychology that is similarly concerned with intelligence at the molar level. The knee-jerk presumption has always been that any such reduction, to the underlying laws of any *one* of the many possible material substrates, would be hopelessly *chauvinistic,* in that it would automatically preclude the legitimate ascription, of the cognitive vocabulary being reduced, to entities composed of any of the many *other* possible material substrates. But this inference needs to be reexamined. It is, in fact, wildly fallacious.

What fuels the inference is the assumption that different material substrates—such as mammalian biology, invertebrate biology, extraterrestrial biology, semiconductor electronics, interferometric photonics, computational hydrology, and so on—will be governed by *different* families of physical laws. But this need not be so. Let me illustrate with three salient and instructive examples.

Sound is a molar-level phenomenon. That is to say, it can be displayed only where there exist a large number of microscopic particles interacting in certain ways. And it, too, is a phenomenon that is multiply realized: in the Earth's highly peculiar atmosphere, in a gas of any molecular constitution, in a liquid of any molecular constitution, and in a solid of any molecular constitution. Sound propagates in any and all of these media. And yet sound is identical with, is smoothly reducible to, compression waves as propagated in any of these highly diverse media. For the underlying physical laws that bring the phenomenon of sound into the embrace of mechanical phenomena generally are *indifferent* to the peculiar molecules that make up the conducting medium, and to their collective status as a gas, liquid, or solid. What matters is that, collectively those particles form an *elastic medium* that allows energy to be transmitted over long distances while the elements of the transmitting medium merely oscillate back and forth a comparatively tiny distance in the direction of energy transmission. To put it bluntly, the very *same* laws of wave propagation in an elastic medium cover *all* of the diverse cases at issue. Idiosyncratic features such as the *velocity* of wave propagation may indeed depend upon the details of the conducting medium (such as the mass of its molecules, and whether they form a gas, liquid, or solid). But the various high-level laws of acoustics (such as $v = \lambda\omega$, and other laws concerning the reflective and refractive behaviors of sound generally) reduce to the very same mechanical laws in all of these diverse cases. A diversity of material substrates here does *not* entail diversity in the underlying laws that govern those diverse substrates. Accordingly, acoustics is not an "autonomous science," devoted to finding laws and ontological categories at its "own level of description." It is but one chapter in the broader mechanics of elastic media.

---

[5] Cf. Fodor, "The Special Sciences," *Synthese,* XXVIII (1974): 77–115.

Temperature, also, is a molar-level phenomenon. And it, too, is a phenomenon that is multiply realized: in the Earth's atmosphere, or in any atmosphere, or indeed, in a gas of any molecular constitution whatever, either pure or mixed. For the temperature of a gas is identical with, is reducible to, the mean level of kinetic energy of the molecules that make up that gas. Here again, the underlying laws of motion (Newton's laws) that govern the behavior of, and the interactions of, the molecules involved are *the very same* for every kind of molecule that might be involved. Those laws are simply indifferent to the shape, or the mass, or the chemical make-up of whatever molecules happen to constitute the gas in question. Idiosyncratic details, such as the velocity of *dispersion* of an unconfined gas, will indeed depend on such details as molecular mass. But the laws of classical thermodymanics (such as the Ideal Gas Law, $PV = \mu RT$) reduce to the same set of underlying mechanical laws whatever the molecular make-up of the gas in question. Once again, a diversity of material substrates does *not* entail diversity in the underlying laws that govern those diverse substances. Accordingly, classical thermodynamics is not an "autonomous science," devoted to finding laws and ontological categories at its "own level of description." Its reduction to statistical mechanics is a staple of undergraduate physics texts.

For a third example, a dipole magnetic field—as instanced in the simple rectangular bar-magnet that one uses to pick up scattered thumb-tacks—constitutes a molar-level phenomenon, but such dipole magnetic fields are realizable in a variety of distinct metals and materials. Pure iron is the most familiar substrate, but sundry alloys (such as aluminum + nickel + cobalt) will also support such a field, as will certain metal/ceramic mixtures. Indeed, *any* substrate that somehow involves charged particles moving in mutually aligned circles (such as a tightly wound current-carrying coil of copper wire) will support a dipole magnetic field. For the simple laws that describe the shape and causal properties of such a field are all reducible to lower-level laws (Maxwell's Equations) that describe the induction of electric and magnetic fields by the motion of charged particles such as electrons. And those lower-level laws are, once again, indifferent to the details of whatever material substrate happens to sustain the circular motion of charged particles.

Once again, an open-ended diversity of sustaining substrates does not entail the irreducibility of the molar-level phenomenon therein sustained. And the historical pursuit of the various pre-Maxwellian theories of dipole magnetic fields (for example, "effluvium" theories) did not constitute an "autonomous science," forever safe from the reductive reach of new and more comprehensive theories. On the contrary, the work of Faraday and Maxwell brought those older theories into the welcoming embrace of the new, and much to the illumination of the former.

These examples can be multiplied. But even one such example illustrates the hazards of inferring the irreducibility of some molar-level phenomenon from the premise of its multiple realizability, even when that premise happens to be *true*. For diverse material substrates may still be governed, all of them, by some of the very *same* low-level physical laws, laws quite capable of explaining the molar-level behaviors *shared* by all of those diverse substrates. The classical argument for element (5) of the functionalist manifesto is plainly flawed, for it naively ignores these obvious sorts of reductive possibilities.

Let us push this line of thought a little farther. For it now begins to appear that, beyond merely providing decisive counterexamples to what Functionalism presents as a robustly *non*reductive pattern, the three examples just cited also provide an *alternative* pattern—a pattern whereby molar-level theories that record genuine high-level regularities across diverse material substrates will *typically* find a successful reduction to some underlying and highly general physical laws, laws that are simply blind to the idiosyncratic and irrelevant differences that happen to distinguish the several substrates. We have just seen this happen in three unproblematic cases. And there are, I repeat, many more.[6]

---

[6] For a fourth example, consider the shared molar-level thermodynamic profile of living organisms across a wide variety of biochemical substrates. The underlying laws of nonequilibrium thermodynamics are once again blind to the peculiar chemical makeup of such diverse substrates. For a fifth example, consider Kepler's three laws of planetary motion, valid for planets of highly diverse material constitution. All three of those laws are reducible to Newton's particle mechanics plus his Universal Law of Gravitation. For a sixth example, consider the science of aerodynamics: the theory of creatures or machines that are capable of flight. Multiple realizability is an obvious feature of this domain: think of seagulls, hummingbirds, bats, dragonflies, wooden airplanes, metal airplanes, helicopters, and so on. And yet their shared molar behavior is ultimately owed to the fact that they all contrive to accelerate ambient air

On this alternative logical and historical pattern, legitimate molar-level theories that comprehend genuine natural kinds will thus be positively *expected* to find some such intertheoretic reduction. For if they eventually prove *not* to be thus reducible, we will have to reconsider the initial presumption that the molar-level theory really does embrace genuine high-level natural kinds governed by genuine high-level explanatory laws. The "unitary" account that the molar theory seemed to provide, across the diverse substrates, might then have to be judged an accidental or a false unity. And its supposedly law-like generalizations will thus turn out to be *accidental* generalizations of some sort, generalizations that are empty of real explanatory and predictive power. Accordingly, if we expect our beloved Folk Psychology, or *any* Psychology, to provide an accurate, natural-kind-embracing, genuinely nomological and explanatory account of the molar-level cognitive operations and behavior of humans, other mammals, human-like aliens, and human-like artificial automata, then we had better *hope* that there exist highly general underlying laws—laws blind to the material differences between all of these diverse creatures—which serve collectively to explain, and thus to reduce, the categories and laws of Psychology.

Let us finally confront the most important question here at issue. Just what are the chances of finding some substrate-neutral underlying laws—laws with a suitably broad explanatory reach—for *Psychology* in particular? That is to say, what are the chances that the case of Psychology will turn out to be an instance of the alternative and overtly *reductive* pattern of development explored in the preceding pages, and in the examples of footnote 6? Certainly they are not zero. For

---

more-or-less continuously downwards, which activity yields, by Newton's (substrate-neutral) third law, a reactive upwards force that is more-or-less continuously equal to the task of keeping them aloft. For a seventh example, consider the closely similar chemical and electrical behaviors of the distinct elements within a given chemical 'family,' those that constitute one vertical column of the Periodic Table (for example, the metals, or, the noble gases). Here the shared molar-level chemical regularities, across a given family of elements, are explained in terms of shared valence-electron-shell structures across the distinct types of atoms within that family. For an eighth example, one rather closer to the case of cognitive creatures, consider the molar behavior of any radio, television, or music player. Despite the great variety of metal and semiconductor substrates that will instantiate the required circuits for signal detection, amplification, and presentation, the behavior of all such devices is reducible to the same set of electrodynamical laws concerning resistances, capacitances, and inductances, laws blind to the material diversity of the substrates that a given manufacturer may choose to employ. As we see from such examples, this general *reductive* pattern, across substrate diversity, is quite robust. For an illuminating discussion, see Michael Strevens, *Bigger than Chaos* (Cambridge: Harvard, 2003), especially chapter 5, "Implications for the Philosophy of the Higher-Level Sciences."

there are at least two low-level theories that have sufficient generality to embrace all of the diverse material realizations of cognitive activity listed in the preceding paragraph, and that also hold promise for explaining at least some of the activities comprehended by Psychology. Let us take a look at them.

V. SOME REDUCTIVE POSSIBILITIES FOR MOLAR-LEVEL PSYCHOLOGY

The first possible framework equal to the task of comprehending all of the diverse material realizations envisioned for cognitive systems is *nonequilibrium* thermodynamics. Distinct from the more familiar (near-equilibrium) statistical thermodynamics discussed earlier, this is the general framework for describing the laws of energy and information flow in partially-closed physical systems that are, and remain, very far away from energetic equilibrium. This is the framework, still in its developmental infancy, that already unites and illuminates all *biological* phenomena, whatever their physical constitution. The basic idea, first outlined half a century ago by the physicist, Erwin Schrodinger,[7] is that any living organism is a highly improbable physical structure whose natural behavior—if it is located in a suitable flow of ambient energy—serves to exploit whatever structure it already contains so as to produce *additional* physical structure. It grows, or it repairs itself, or it reproduces. Such an interest-bearing investment[8] is possible only when the system is situated so as to exploit an energy flux that begins with energy from a very low-entropy state,[9] energy that is then progressively dissipated into energy at a much higher entropy state. The living physical system 'steals' some of the initial low entropy energy as that energy courses through it, and it then incorporates that energy in the form of additional (and improbable) physical structure. The low-entropy energy source for our terrestrial environment is ultimately the Sun, radiating at a black-body temperature of roughly 4000°K (at rather short wavelengths). And the ultimate high-entropy energy sink is the surrounding background of empty space, radiating at a black-body temperature of about 3°K (at very long wavelengths). In between lies the biosphere at a temperature around 293°K. Without such a concentrated or low-entropy energy source "above" us, and such a dissipated high-entropy energy sink

---

[7] *What Is Life?* (Cambridge: Harvard, 1944).

[8] Note well the economic metaphor here deployed. Its aptness will come up again shortly.

[9] Entropy is a measure of how chaotically *dissipated* an amount of energy happens to be, a measure of how *unavailable* it is to do work, a measure (if you like) of its *weariness*. By contrast, a low-entropy state implies energy that is highly "concentrated" and available to do work.

"below" us, nothing alive could hope to remain alive. Indeed, nothing of any biological interest—that is, no extremely improbable physical structures with complex metabolic pathways—could ever have evolved in the first place.

Those that have evolved are thus instant testament to the existence of such a complexity-inducing ambient energetic waterfall—a constant flow from the sun, through us, and into the cold abyss beyond. Moreover, any individual of any species also embodies, in its typical structural details, extraordinary amounts of information about the peculiar environmental niche in which it thrives. For no individual could be expected to *have* the specific physical structure it has *unless* the environment in which it thrives has a comparably specific physical and dynamical profile. For, once again, eons of evolutionary pressures have made the former exquisitely "tuned" to the latter, functionally and metabolically speaking. If you want an indirect but highly informative window onto the chemical and biological dynamics of a forest pond, examine the frog who lives there.

I provide the reader with this brief sketch, of how nonequilibrium thermodynamics embraces biological phenomena generally, not just to validate the claim made in footnote 6, but to prepare the ground for the following suggestion. Cognitive phenomena are just an additional instance or iteration of the thermodynamical profile already outlined. Specifically, a creature that *learns* about the world is a creature that exploits the low-entropy internal structure or information that it already possesses, in such a fashion that, if the creature is placed in an environment with an information-rich energy flow, it comes to embody *additional* information-bearing structure about its environment, typically in the form of a progressively re-wired brain.

The relevant energy flow here begins with the low-entropy states of incoming *sensory* signals (light, sound, pressure, taste, smell, whatever), signals that contain detailed information about the immediate physical environment. And it ends with the dissipation of that original low-entropy energy, after the brain's cognitive processing is done, as high-entropy heat radiated away by the body at long wavelengths in the infrared, its original information "lost," or rather, left behind in the brain. The active cognitive system "steals" some of the low-entropy energy that its sensory organs provide, and incorporates it as additional information-bearing structure. My biological body at age six days will embody a great detail of general information about my natural environment, as we noted two paragraphs ago, concerning the frog. But my brain, at six years (or at six decades), embodies an additional wealth of information, information that the human genome was, and is, far too poor to bequeath to me. I have to acquire that wealth of information post-natally. And I do. And so does a mouse. And so does any cognitive creature. For that is what a cognitive creature is: an "extra-somatic information multiplier." Unlike a typical Heat Engine (for example, a steam engine, an automobile engine), which exploits an entropy-increasing energy flow to produce macroscopic *motion*, the brain exploits such a flow to produce neuronally and synaptically-embodied *information*. We are, in fact, *Epistemic* Engines, not just figuratively, but literally.

This naturalistic portrait, note well, makes no reference to the variety of material substrates that might sustain such an energetic and informational economy. Many different substrates are presumably possible. What makes them all cognitive creatures—part of it, anyway—is their shared thermodynamic and information-multiplying profile.

I briefly floated this possible construal of cognitive creatures some twenty-five years ago, in the closing paragraphs of my 1979 book,[10] and again, in slightly more detail, in a 1982 paper.[11] Those accounts of cognition, and that of the preceding two paragraphs, may well be dismissed as mere handwaving speculation, unless we can provide an account of *how* brains actually process, and incorporate into their internal structure, ambient information.

In the salient case of *biological* metabolisms, we do indeed possess such a nonhandwaving account. We know how DNA embodies information. We know how that information is read out by RNA in order to synthesize various protein molecules. We know how those protein molecules catalyze certain metabolic reactions, and sequences of such reactions. We know how those reaction chains create new biological molecules that form additional biological structures. We know how those structures collectively steer ambient energy and materials along paths that sustain and amplify the organism at issue. The nonequilibrium thermodynamical portrait of living things is therefore not just a philosophical guess. It is a highly general reductive framework that brings real illumination to biological processes, across a wide diversity of chemical substrates.

Twenty-five years ago, I must own, the nonequilibrium thermodynamical portrait of *cognitive* activity *was* a merely philosophical guess—a hesitant extrapolation from the thermodynamic portrait of living things just explored. For we then lacked any corresponding account of how brains actually process and eventually store new information. In the intervening years, however, an account of exactly those activities has pieced itself together, and has become the focus of a great deal of research, both experimental and theoretical. That account posits fleeting *activation vectors* across proprietary populations of neurons as the basic mode of ephemeral representation. It posits *vector-to-vector transformations*, at the hands of intervening matrices of synaptic connections, as the basic mode of computation over those representations. It posits specific *configurations of weighted synaptic connections* as the basic mode of general or background knowledge. And it posits ongoing *adjustments in the values* of those weighted synaptic connections as the most basic mode of learning.[12]

---

[10] *Scientific Realism and the Plasticity of Mind* (Cambridge: Harvard, 1979).

[11] "Is *Thinker* a Natural Kind?" *Dialogue*, XXI, 2 (1982): 223–38.

It is vital to appreciate that the structural and dynamical portrait just painted—of vector coding and vector processing via large matrices with plastic coefficients—is once again a portrait that can be realized in a wide variety of material substrates: in mammalian brains, in octopus brains, in extraterrestrial brains, in electronic chips, in optical systems, and so forth. For the mathematical laws of vector/matrix processing, and of the information-dependent, experience-driven adjustments of the individual coefficients of the transforming matrix, are all indifferent to the physical medium that displays those fleeting vectors and embodies those comparatively enduring matrices. Some idiosyncratic details, such as cognitive reaction times, will indeed be sensitive to the implementational facts, such as the speed of vectorial conduction between distinct populations of active units. (As noted earlier, an electric current in a copper wire propagates much faster than a spike train in an axon.) But the profile of molar-level activity will be importantly similar across all of these diverse substrates.

Once again we are contemplating a low-level explanation, in terms of general or abstract underlying natural laws, of a roughly constant profile of molar-level activity, activity that can be displayed across a considerable variety of material substrates. But this time the explanatory target is the *cognitive* activities of creatures like ourselves. We are no longer pressing a mere a priori possibility on our functionalist friends. We are confronting a pair of real theories (the vector/matrix account of brain structure and function, and the nonequilibrium thermodynamical account of brain growth and learning), theories that hold some nontrivial promise of providing systematic reductive explanations of cognitive phenomena in particular, despite their presumably diverse realizations. To put it bluntly, we are confronting exactly what classical functionalism said was not to be had, nor even to be sought.

---

[12] For accessible, entry-level accounts that will provide doors onto the wider literature, see Patricia S. Churchland and Terrence J. Sejnowski, *The Computational Brain* (Cambridge: MIT, 1987); my *The Engine of Reason, The Seat of The Soul* (Cambridge: MIT, 1995); and *Inner Spaces and Outer Spaces: The New Epistemology*.

These new developments, especially the vector/matrix story, have already given us a much deeper understanding of what the brain does, and of how it manages to do it. We now understand, for example, how the activation space of a large population of neurons can come to embody a structured system of categorical prototypes—that is, a meaningful *conceptual framework*. We now understand how those prototype points in such a background neuronal activation space can be selectively activated by sensory inputs. That is, we understand how a brain can *interpret* its sensory experience in terms of its acquired conceptual framework. We now understand how prototypical motor behaviors can be represented as prototypical activation-trajectories in motor-neuron activation space. That is, we have some understanding of how complex *motor skills* are embodied. And we know how such unfolding trajectories can actually *generate* the relevant motor behaviors in the body's limb and muscle systems. In sum, we can now see cognitive activity as we have never seen it before. Whether we are seeing it correctly, only time will tell. But a fertile vision is already being explored. What this means is that the celebrated element (5) of the functionalist manifesto is not just naively argued. In fact, it is almost certainly false.[13]

IV. WHAT DOES *NOT* UNITE THE CLASS OF COGNITIVE CREATURES

On the vector/matrix story explored above, what carries the burden of any creature's acquired background knowledge, of the world's general and enduring structure, is the specific configuration of the billions or trillions of synaptic connections that variously intervene between the brain's many distinct neuronal coding populations. It is these variously weighted excitatory or inhibitory elements that constitute the brain's principal memory store, and also its principal means of computation. One and the same system simultaneously serves both functions.

---

[13] Allow me a closing remark on Fodor's 1974 parade case of a molar-level natural science for which reductive aspirations are supposed to be clearly foolish, namely Economics. The supporting argument then appealed to the multiple realizability of *currency* systems—such as shell currency, coin currency, paper currency, electronic currency, and so on. We can all agree that Economics is not going to be reducible to the chemistry of wood fiber, or to the physics of copper and gold. But all of this is now visibly beside the point. For we can now appreciate that Economics is the study of the *metabolisms of superorganisms*, a phenomenon that once again falls firmly within the province of nonequilibrium thermodynamics, a science whose laws are blind to such implementational details. A national economy, after all, embodies a flow of both energy and materials: it creates real physical and organizational structures, and it dissipates vast amounts of (initially low-entropy) energy in the process. It is too soon to *insist* that Economics will indeed find such an explanatory reduction. But neither can Fodor justly insist that it will not. The presumption in favor of his principal nonreductive example has just evaporated.

No two people, however, display the same configuration of synaptic connections and synaptic weights. Each human brain boasts roughly 10^{14} synaptic connections, the overwhelming majority of which are established post-natally in response to a lived experience that is unique to each individual. Since we experience a common world that does display enduring features, each of us ends up with a family of sculpted activation spaces whose structure is *similar* to the structure of other people's activation spaces, at least if they are members of the same culture. But genuine identity is too much to ask for. We may all agree that the vector/matrix system found in each individual is computing a function of some fabulous complexity. But no two people on the planet will be computing exactly the same function, for no two people share the same matrix of synaptic connections.

Very well, but surely they will be computing *similar* functions? Indeed so, if they happen to be peas from the same cultural pod.[14] But what needs emphasizing here is the real functional *diversity* displayed by individuals at different stages of their lives, in different cultures across the planet, and at different points in our very long and cognitively diverse human history. This diversity in the functions that brains are computing becomes more striking still when we expand our consideration to include cognitive activity in the nervous systems of other terrestrial creatures such as chimpanzees, cats, mice, finches, crocodiles, crabs, octopi, fish, and spiders. Clearly such diverse creatures are not all computing the same function, nor even remotely similar functions. And yet, we are all cognitive creatures.

What is it, then, that unites us? Ironically, it appears to be the abstract form of our *hardware* that unites us! We are, all of us on the preceding list, massively parallel vector-processors whose ever active vector-transforming matrices (our trillions of synaptic connections) are slowly updated or instructed by a procedure that filters information from a low-entropy flux of energy from our sensory peripheries. This computational arrangement has prodigious advantages over the serial architecture deployed in classical (von Neumann) computers—in its speed of computation, in its graceful tolerance of scattered component failures, and in its swift deployment of relevant information. This alternative computational template is sufficiently virtuous to make it a likely evolutionary attractor on any planet that develops life, not just on Earth, and to make it a compelling technological choice for any future attempts at constructing artificial intelligence as well.

---

[14] How this similarity can be achieved, despite our synaptic diversity, is detailed in my "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered," this JOURNAL, XCV, 1 (January 1998): 5–32.

Accordingly, element (1) of the original functionalist manifesto is almost certainly a mistake—indeed, a monumental mistake. The cognitive creatures on this planet are computing a bewildering *variety* of very different functions, but they are all using fundamentally similar computational "hardwares" in order to do it. On this fundamental point, classical functionalism had things exactly backwards.

Element (2) must therefore be rejected as well. If our alternative portrait of cognition is even roughly correct, the central job of Cognitive Psychology is to explore *how* it is that terrestrial brains are able to compute the extraordinary *variety* of functions displayed in diverse species of cognitive creature. This must be an empirical undertaking, one sensitive to the idiosyncrasies of nonhuman nervous systems. Accordingly, element (3) must be rethought along the same lines. The central job of AI research is not just to explore the construction of artificial vector-processing systems that compute the same function that some species of animal is already computing. A central part of its job will be to explore instead the pregnant potential of such artificial systems for computing functions—for pursuing cognitive activities—that no terrestrial creature has yet pursued or ever will pursue. Large-scale electronic realizations of our vector/matrix-style of computational resources will explore entirely new horizons for information processing and world representation, and, being electronic, they will do it roughly a million times faster than biological creatures can ever hope to do it (because the speed of signal conduction in a copper wire is close to the speed of light). The enterprise of Artificial Intelligence thus has a dazzling future, but not because classical functionalism launched it in the right direction.

Indeed, it launched the enterprise in a most unfortunate direction. Element (6) provided a twisted rationale for mostly *ignoring* the empirical or experimental neurosciences, and for ignoring the early theoretical work that attempted to model the activities of large numbers of interconnected neurons. Worse still, element (7) celebrated this deliberate disconnection with an ill-conceived positive portrait of Cognitive Psychology and Artificial Intelligence as "methodologically autonomous sciences." In retrospect, this was unwise, despite the genuinely clever contributions of a great many gifted researchers. For it served to insulate the relevant research from exactly the empirical information that promised the most interesting and authoritative constraints on whatever models were put up for evaluation. The result was almost half a century of misdirected research.

V. CONCLUSIONS

The seven elements of our opening functionalist manifesto, all seven of them, appear to be false—not just inadequately argued for, but outright false. Fortunately, we now have an alternative positive portrait of cognitive activity in place, one capable of steering systematic research on the nature of cognition, research that points in directions interestingly different from the directions that dominated the last half of the twentieth century. That alternative portrait is no longer imprisoned by the linguaformal representational paradigm provided by Folk Psychology, nor is it centered on the computational paradigm provided by the digital-coding, serial-processing, program-running machines of the still-standard von Neumann configuration. Instead, it draws its opening inspirations, in both cases, from the deeply instructive empirical example of terrestrial nervous systems.

In closing, it is worth pointing out that two prominent background assumptions of the functionalist program have not been denied in the preceding critique. The first assumption is that cognitive creatures are indeed engaged in computing complex functions of some sort or other. And the second is that these computational activities, whatever they are, can be realized in a diversity of physical substrates. These assumptions are presumably as true, and as important, as they ever were. But in the present intellectual environment, those same two assumptions now pull our imaginations in entirely new directions. The first assumption motivates the brain-centered research program known as Computational Neurobiology. And the second assumption motivates the development of alternative physical realizations (presumably electronic or photonic)—not of our "software" (strictly speaking, we do not *have* any!), but—of the massively parallel, vector-processing structure of our biological *hardware*. Let us hope that this second wave of research will be more revealing, and less self-blinkered, than the functionalist-inspired wave that preceded it.

PAUL M. CHURCHLAND

University of California, San Diego

MEMO TO AUTHORS

The Journal of Philosophy is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally publish to no more than 7,500 words, or 18 *Journal* pages. A more precise way of determining length is by characters (including spaces): 2,850 characters per *Journal* page, or 51,300 characters per *Journal* article of 18 pages.

The editors request that three copies of all manuscripts submitted be prepared according to the following guidelines:

(1) All material, including the quotations and footnotes, must be typed, double-spaced, and one-sided, with ample margins left and right, top and bottom.
(2) No bibliographies will be published; bibliographical material should be put into footnotes.
(3) Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include: name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include: publisher, U.S. city of publication, and date. Footnotes should be gathered at the end of the article.
(4) Quotations should be checked by the author; capitalization, italics, spelling, and the like should match the original.
(5) If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked. Standard logic symbols are available, and letters of the Greek and script alphabet.
(6) No underlining; please limit use of italics to technical terms (when they are introduced) and to variables.
(7) No contractions.

The editors are willing to read and evaluate a manuscript that does not meet these guidelines, provided the author agrees to meet them before submitting a final manuscript. The editors may ask for revision of excessively long manuscripts, and may require reformatting of accepted manuscripts. They also request an electronic version with the final hard copy.

The Journal of Philosophy does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere. While we do not publish manuscripts that mainly comment on articles published elsewhere, we do have a "Comments and Criticism" section in which short responses to *Journal* articles may be published.

Manuscripts may be sent to: John Smylie, Managing Editor, *The Journal of Philosophy*, 709 Philosophy Hall, 1150 Amsterdam Avenue, Mail Code 4972, New York, NY 10027, Email jys3@columbia.edu.

NEW BOOKS

FLORENCE, PENNY: *Sexed Universals in Contemporary Art.* New York: Allworth, 2004. xvii, 190 p. Paper $19.95.

FLYNN, THOMAS R.: *Sartre, Foucault, and Historical Reason, Volume Two.* Chicago: University Press, 2005. xviii, 390 p. Cloth $62.50, paper $25.00.

FORSTER, MICHAEL N.: *Wittgenstein on the Arbitrariness of Grammar.* Princeton: University Press, 2004. xiii, 245 p. Cloth $29.95.

FRANCO, PAUL: *Michael Oakeshott: An Introduction.* New Haven: Yale, 2004. xi, 209 p. Cloth $30.00.

FRANK, JILL: *A Democracy of Distinction: Aristotle and the Work of Politics.* Chicago: University Press, 2005. xiii, 199 p. Cloth $49.00, paper $19.00.

FRANKLIN, JULIAN H.: *Animal Rights and Moral Philosophy.* New York: Columbia, 2005. xix, 151 p. Cloth $35.00.

FRIEDLANDER, ELI: *J.J. Rousseau: An Afterlife of Words.* Cambridge: Harvard, 2004. viii, 160 p. Cloth $39.95.

FULLER, STEVE: *Kuhn vs. Popper: The Struggle for the Soul of Science.* New York: Columbia, 2004. ix, 143 p. Cloth $24.50.

FYNSK, CHRISTOPHER: *The Claim of Language.* Minneapolis: Minnesota UP, 2004. xvi, 106 p. Cloth $53.95, paper $17.95.

GALE, RICHARD M.: *The Philosophy of William James: An Introduction.* New York: Cambridge, 2005. x, 246 p. Cloth $65.00, paper $23.99.

GANE, NICHOLAS: *The Future of Social Theory.* New York: Continuum, 2004. xii, 210 p. Cloth $105.00, paper $27.95.

GILROY, PAUL: *Postcolonial Melancholia.* New York: Columbia, 2005. xvi, 170 p. Cloth $24.50.

GOLASH, DEIRDRE: *The Case Against Punishment.* New York: University Press, 2005. ix, 217 p. Cloth $45.00.

GOLDSTEIN, PHILLIP: *Post-Marxist Theory: An Introduction.* Albany: SUNY, 2005. vii, 143 p. Cloth $57.50, paper $16.95.

GORDON, HAIM, and SHLOMIT TAMARI: *Maurice Merleau-Ponty's Phenomenology of Perception: A Basis for Sharing the Earth.* Westport, CT: Praeger, 2004. 148 p. Cloth $79.95.

GORMAN, JONATHAN: *Rights and Reason: An Introduction to the Philosophy of Rights.* Ithaca, NY: McGill-Queen's, 2003. x, 230 p. Cloth $70.00, paper $27.95.

GRAHAM, GORDON: *Eight Theories of Ethics.* New York: Routledge, 2004. x, 221 p. Cloth $80.00, paper $22.95.

GROSZ, ELIZABETH: *The Nick of Time: Politics, Evolution, and the Untimely.* Durham, NC: Duke, 2004. ix, 314 p. Paper $22.95.

GUTTING, GARY: *Continental Philosophy of Science.* Malden, MA: Blackwell, 2005. xii, 332 p. Cloth $69.95, paper $34.95.

HARMAN, GRAHAM: *Tool-Being: Heidegger and the Metaphysics of Objects.* Chicago: Open Court, 2002. vii, 331 p. Paper $29.95.

HART, KEVIN: *The Dark Gaze: Maurice Blanchot and the Sacred.* Chicago: University Press, 2004. x, 301 p. Cloth $55.00, paper $22.00.

HEBBLETHWAITE, BRIAN: *Philosophical Theology and Christian Doctrine.* Malden, MA: Blackwell, 2004. xi, 176 p. Cloth $64.95, paper $29.95.

HELLER-ROAZEN, DANIEL: *Echolalias: On the Forgetting of Language.* New York: Zone, 2005. 273 p. Cloth $28.00.

HENDRICKS, VINCENT F.: *Feisty Fragments for Philosophy.* London: King's College, 2004. xv, 188 p. Paper $24.00.

HIRSTEIN, WILLIAM: *Brain Fiction: Self-Deception and the Riddle of Confabulation.* Cambridge: MIT, 2005. ix, 289 p. Cloth $35.00.

HUTCHENS, B.C.: *Levinas: A Guide for the Perplexed.* New York: Continuum, 2004. viii, 191 p. Cloth $85.00, paper, $21.95.

JAMES, MICHAEL RABINDER: *Deliberative Democracy and the Plural Polity.* Lawrence: Kansas UP, 2004. xii, 239 p. Cloth $35.00, paper $17.95.

JAY, MARTIN: *Songs of Experience.* Berkeley: California UP, 2005. x, 431 p. Cloth $34.95.

JOHNSON, DAVID: *Truth without Paradox.* Lanham, MD: Rowman and Littlefield, 2004. xi, 195 p. Cloth $70.00.

KAINZ, HOWARD P.: *Natural Law: An Introduction and Re-examination.* Chicago: Open Court, 2004. xvi, 151 p. Paper $25.95.

KAMBHAMPATI, UMA S.: *Development and the Developing World.* Malden, MA: Polity, 2004. x, 296 p. Paper $29.95.

KAVKA, MARTIN: *Jewish Messianism and the History of Philosophy.* New York: Cambridge, 2004. xiii, 241 p. Cloth $65.00.

KENNY, ANTHONY: *Ancient Philosophy: A New History of Western Philosophy, Volume 1.* New York: Oxford, 2004. xxi, 341 p. Cloth $30.00.

KETCHAM, RALPH: *The Idea of Democracy in the Modern Era.* Lawrence: Kansas UP, 2004. viii, 302 p. Cloth $35.00.

KIERKEGAARD, SØREN: *The Humor of Kierkegaard.* THOMAS C. ODEN, ed. Princeton: University Press, 2004. x, 338 p. Cloth $55.00, paper $16.95.

KIRWIN, JAMES: *The Aesthetic in Kant: A Critique.* New York: Continuum, 2004. vi, 200 p. Cloth $105.00.

LEVENE, NANCY K.: *Spinoza's Revelation: Religion, Democracy, and Reason.* New York: Cambridge, 2004. xxi, 256 p. Cloth $75.00.

LIVINGSTON, PAUL M.: *Philosophical History and the Problem of Consciousness.* New York: Cambridge, 2004. xiii, 279 p. Cloth $70.00.

LYNCH, MICHAEL P.: *True to Life: Why Truth Matters.* Cambridge: MIT, 2004. xii, 204 p. Cloth $27.95.

MALLE, BERTRAM F.: *How the Mind Explains Behavior.* Cambridge: MIT, 2004. viii, 314 p. Cloth $38.00.

MARGOLIS, JOSEPH: *Moral Philosophy after 9/11.* University Park: Pennsylvania UP, 2004. xvii, 150 p. Cloth $29.95.

MCGINN, COLIN: *Consciousness and Its Objects.* New York: Oxford, 2004. ii, 256 p. Cloth $35.00.

MELSER, DEREK: *The Act of Thinking.* Cambridge: MIT, 2004. x, 287 p. Cloth $35.00.

MILLIKAN, RUTH GARRETT: *Varieties of Meaning: The 2002 Jean Nicod Lectures.* Cambridge: MIT, 2004. xi, 242 p. Cloth, $35.00.

MORAVCSIK, JULIUS M.: *The Ties That Bind.* New York: Central European UP, 2004. xiii, 294 p. Cloth $47.95.

NOË, ALVA: *Action in Perception.* Cambridge: MIT, 2004. xi, 277 p. Cloth $38.00.

OAKLANDER, L. NATHAN: *The Ontology of Time.* New York: Prometheus, 2004. 366 p. Paper $30.00.

O'CONNOR, BRIAN: *Adorno's Negative Dialectic: Philosophy and the Possibility of Critical Rationality.* Cambridge: MIT, 2004. xviii, 224 p. Cloth $35.00.

PORTMANN, JOHN: *Bad for Us: The Lure of Self-Harm.* Boston: Beacon, 2004. xxii, 224 p. Cloth $25.00.

ROBINSON, DAVID M.: *Natural Life: Thoreau's Worldly Transcendentalism.* Ithaca, NY: Cornell, 2004. xiv, 234 p. Cloth $24.95.

ROOCHNIK, DAVID: *Retrieving the Ancients: An Introduction to Greek Philosophy.* Malden, MA: Blackwell, 2004. vii, 238 p. Paper $22.95.

ROW, WILLIAM L.: *Can God Be Free?* New York: Oxford, 2004. vii, 173 p. Cloth $45.00.

ROWLANDS, MARK: *The Philosopher at the End of the Universe: Philosophy Explained through Science Fiction Films.* New York: St. Martin's, 2004. xi, 276 p. Cloth $23.95.

ROY, M.N.: *Radical Humanist: Selected Writings.* INNAIAH NARISETTI, ed. Amherst, NY: Prometheus, 2004. 209 p. Cloth $28.00.

RUSSON, JOHN: *Reading Hegel's Phenomenology.* Bloomington: Indiana UP, 2004. xi, 299 p. Cloth $50.00, paper $27.95.

SCHMAUS, WARREN: *Rethinking Durkheim and His Tradition.* New York: Cambridge, 2004. xii, 195 p. Cloth $65.00.

SCHROEDER, WILLIAM R.: *Continental Philosophy: A Critical Approach.* Malden, MA: Blackwell, 2004. xviii, 456 p. Cloth $64.95, paper $29.95.

SEARLE, JOHN R.: *Mind: A Brief Introduction.* New York: Oxford, 2004. viii, 326 p. Cloth $26.00.

SEDLEY, DAVID: *The Midwife of Platonism: Text and Subtext in Plato's "Theaetetus."* New York: Oxford, 2004. x, 201 p. Cloth $60.00.

SMITH, MICHAEL: *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics.* New York: Cambridge, 2004. xii, 388 p. Cloth $80.00, paper $32.99.

SPARGO, R. CLIFTON: *The Ethics of Mourning: Grief and Responsibility in Elegiac Literature.* Baltimore: Johns Hopkins, 2004. x, 314 p. Cloth $39.95.

SPARIOSU, MIHAI I.: *Global Intelligence and Human Development.* Cambridge: MIT, 2004. xi, 274 p. Cloth $57.00, paper $23.00.

STERBA, JAMES P.: *The Triumph of Practice over Theory in Ethics.* New York: Oxford, 2005. xi, 206 p. Cloth $45.00, paper $18.50.

STONE, JON R.: *The Routledge Dictionary of Latin Quotations.* New York: Routledge, 2005. xvii, 394 p. Paper $17.95.

STREVENS, MICHAEL: *Bigger than Chaos: Understanding Complexity through Probability.* Cambridge: Harvard, 2003. xiii, 413 p. Cloth $59.95.

TAN, KOK-CHOR: *Justice without Borders.* New York: Cambridge, 2004. xiii, 219 p. Cloth $70.00, paper $25.99.

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2005

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new option, click on the PayPal icon at
www.jphil.org

new from The MIT Press

True to Life

Why Truth Matters

Michael P. Lynch

"Michael Lynch explains with engaging energy and clarity why the concept of truth matters to a decent public culture." — Martha Nussbaum, The University of Chicago

A Bradford Book • 224 pp. $27.95

Is Oedipus Online?

Siting Freud after Freud

Jerry Aline Flieger

Explores psychoanalysis as a navigation device for the cultural maze of the twenty-first century.

Short Circuits series
428 pp., 43 illus. $19.95 paper

Action in Perception

Alva Noë

"The most lucid and persuasive defense of the enactive theory of perception that I have read."
— C. L. Hardin, Syracuse University

392 pp., 10 illus. $38

There's Something About Mary

Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument

edited by Peter Ludlow,
Yujin Nagasawa, and Daniel Stoljar
foreword by Frank Jackson

Key papers on one of the most important and provocative thought experiments in philosophy of mind.

A Bradford Book • 480 pp., 2 illus. $35 paper

Pragmatism, Critique, Judgment

Essays for Richard J. Bernstein

edited by Seyla Benhabib
and Nancy Fraser

Leading philosophers and social thinkers, including Richard Rorty, Jacques Derrida, and Jürgen Habermas, pay tribute to the influential American philosopher Richard J. Bernstein.

400 pp., 4 illus. $32 paper

Reconstructing Reason and Representation

Murray Clarke

A study of the philosophical implications of evolutionary psychology, suggesting that knowledge is a set of natural kinds housed in the modules of a massively modular mind.

A Bradford Book • 192 pp., 4 illus. $35

Veritas

The Correspondence Theory and Its Critics

Gerald Vision

"Essential reading for those interested in current philosophical debates about truth. Vision lays down the gauntlet against deflationism." — Terry Horgan, University of Arizona

A Bradford Book • 320 pp., 2 illus. $36

Molecular Models of Life

Philosophical Papers on Molecular Biology

Sahotra Sarkar

"Sarkar stands out amongst philosophers of science for his ability to combine conceptual, historical and technical considerations in compelling perspectives on the philosophical questions raised by contemporary biology." — Paul E. Griffiths, University of Pittsburgh

Life and Mind series • A Bradford Book
352 pp. $38

Approaches to Studying World-Situated Language Use

Bridging the Language-as-Product and Language-as-Action Traditions

edited by John C. Trueswell
and Michael K. Tanenhaus

Takes the first steps toward merging the cognitive and social approaches to language processing.

Learning, Development, and Conceptual Change
A Bradford Book • 368 pp., 70 illus. $35 paper

Varieties of Meaning

The 2002 Jean Nicod Lectures

Ruth Garrett Millikan

"An outstanding and exciting book on the relationship between purpose and reference, and essential reading for cognitive scientists interested in the naturalization of intentionality."
— Vittorio Gallese, University of Parma

256 pp., 3 illus. $35

The Act of Thinking

Derek Melser

"An extraordinary work, and well argued too." — William G. Lycan, University of North Carolina at Chapel Hill

304 pp. $35

Brain Fiction

Self-Deception
and the Riddle of Confabulation

William Hirstein

"Both a neuroscientist and a philosopher, William Hirstein writes from his unique vantage point with great scholarship, precision, and clarity to tackle some of the deeper mysteries of the human mind." — Fredric Schiffer, M. D., Harvard Medical School, author of *Of Two Minds*

288 pp., 16 illus. $35

now in paperback

Being No One

The Self-Model Theory
of Subjectivity

Thomas Metzinger

"Metzinger's interdisciplinary approach opens up a new path toward a scientific theory of consciousness and self-consciousness." — *Science*

711 pp., 12 illus. $32 paper

now in paperback

Furnishing the Mind

Concepts
and Their Perceptual Basis

Jesse J. Prinz

"Prinz's discussions . . . provide a clear overview of the field, both in philosophy and psychology." — *Choice*

375 pp., 12 illus. $25 paper

To order call **800-405-1619.**

http://mitpress.mit.edu

Ethics and the Metaphysics of Medicine

Reflections on Health
and Beneficence

Kenneth A. Richman

"Like few authors in the field, Richman situates his bioethics within a broader analysis of issues in the philosophy and metaphysics of medicine."
— H. Tristram Engelhardt, Jr., Professor, Rice University, Professor Emeritus, Baylor College of Medicine

Basic Bioethics series • 232 pp. $29

Is Human Nature Obsolete?

Genetics, Bioengineering, and the Future of the Human Condition

edited by Harold W. Baillie
and Timothy K. Casey

"This excellent, probing collection of papers—spanning philosophy, history, social science, and theology—will be welcomed by professionals and lay readers alike." — Edward Tenner, author of *Why Things Bite Back* and *Our Own Devices*

Basic Bioethics series
440 pp. , 1 illus. $27 paper

Mechanical Bodies, Computational Minds

Artificial Intelligence
from Automata to Cyborgs

edited by Stefano Franchi
and Güven Güzeldere

Considers the past, present, and future of artificial intelligence from a multidisciplinary perspective.

A Bradford Book • 544 pp., 43 illus. $45 paper

Global Intelligence and Human Development

Toward an Ecology
of Global Learning

Mihai I. Spariosu

"A rare flash of intellectual creativity and an important contribution to the study of global transformation."
— Calestous Juma, Harvard University

296 pp. $23 paper

Illuminating.

From a Philosophical Point of View

Selected Studies

Morton White

One of the most important philosophers of recent times, Morton White has spent a career building bridges among the increasingly fragmented worlds of the humanities, social sciences, and natural sciences. *From a Philosophical Point of View* is a selection of White's best essays, written over a period of more than sixty years.

Cloth $45.00 ISBN 0-691-11959-7
Due January

"*From a Philosophical Point of View* has an intellectual unity that results when a single mind of real ability looks about and surveys a wide variety of interesting issues. The author's adroit prose makes it a pleasure to read his discussions."

—Nicholas Rescher,

Reference and Description

The Case against Two-Dimensionalism

Scott Soames

In this book, Scott Soames defends the revolution in philosophy led by Saul Kripke, Hilary Putnam, and David Kaplan against attack from those wishing to revive descriptivism in the philosophy of language, internalism in the philosophy of mind, and conceptualism in the foundations of modality. After explaining the major issues, Soames launches a devastating critique of the two-dimensionalist revival of descriptivism. He then outlines a lasting, nondescriptivist perspective on meaning and a nonconceptualist understanding of modality.

Cloth $39.50 ISBN 0-691-12100-1
Due February

"This outstanding book will define the literature on two-dimensionalism for years to come."

—Ted Sider,
Rutgers University

Important.

Hegel's Preface to the *Phenomenology of Spirit*

Georg Wilhelm Friedrich Hegel

Translation and running commentary by Yirmiyahu Yovel

This is a new translation, with running commentary, of what is perhaps the most important short piece of Hegel's writing. This new translation combines readability with maximum precision, breaking Hegel's long sentences and simplifying their often complex structure. At the same time, it is more faithful to the original than any previous translation.

Cloth $19.95 ISBN 0-691-12052-8
Due February

"Yovel brings impressive exegetical skill and erudition to bear in a way that consistently throws great light on Hegel's text."

—Michael N. Forster,
University of Chicago

Søren Kierkegaard

A Biography

Joakim Garff

Translated by Bruce H. Kirmmse

Joakim Garff's biography of Kierkegaard is a landmark achievement. A seamless blend of history, philosophy, and psychological insight, all conveyed with novelistic verve, this is the most comprehensive and penetrating account yet written of the life and works of the enigmatic Dane who changed the course of intellectual history.

Cloth $35.00 ISBN 0-691-09165-X
Due February

Praise for the original, Danish edition:

"A well thought-out synthesis of Kierkegaard's life and writings so exceptional ... so concrete and rich with perspectives, that it has no equal in literature. Read, read, read."

—*Weekendavisen*

PRINCETON
University Press

800-777-4726 • READ EXCERPTS AT
WWW.PUP.PRINCETON.EDU

Philosophy from Yale

Adorno

A Political Biography

Lorenz Jäger

Translated by Stewart Spencer

This lucid biography of Theodor Adorno, co-founder of the famous Frankfurt School and one of Germany's most important twentieth-century thinkers, offers a fresh perspective on his life and writings. $32.00

Michael Oakeshott

An Introduction

Paul Franco

This authoritative book, which introduces readers to the life and thought of this eminent philosopher, "is both **sympathetic and scrupulously fair**; it is also remarkably user friendly." —Noel Malcolm, *The Sunday Telegraph* $30 00

Symbolic Forms and Cultural Studies

Ernst Cassirer's Theory of Culture

Edited by Cyrus Hamlin and John Michael Krois

"This is **the most comprehensive collection** of serious scholarly studies on Cassirer's thought since the Library of Living Philosophers released its collection over fifty years ago."—Randall E. Auxier, Editor, Library of Living Philosophers

$65 00

A Guide to Philosophy in Six Hours and Fifteen Minutes

Witold Gombrowicz

Translated by Benjamin Ivry

Ann Dropick assisted in preparing the English translation

In this inspired book—a literary gem full of sardonic wit, brilliant insights, and provocative criticism—eminent Polish author Witold Gombrowicz reflects on **Kant, Hegel, Sartre, Schopenhauer, Kierkegaard, and Heidegger.** $22.00

Persuasion and Rhetoric

Carlo Michelstaedter

Translated with an introduction and commentary by Russell Scott Valentino, Cinzia Sartini Blum, and David J. Depew

"Michelstaedter's book is a **still undervalued masterpiece of the XIXth century.**"—Umberto Eco

Italian Literature and Thought $32.50

New in paper

Hannah Arendt

For Love of the World, Second Edition

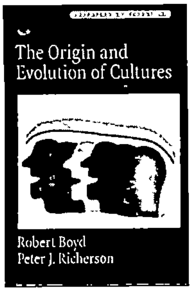
Elisabeth Young-Bruehl

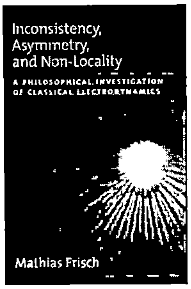
"Both a personal and an intellectual biography . . . **It represents biography at its best.**"—Peter Berger, front page, *The New York Times Book Review* (on the earlier edition)

$22 00 paperback

YALE University Press yalebooks.com

New from
Oxford




Inconsistency, Asymmetry, and Non-Locality
A PHILOSOPHICAL INVESTIGATION OF CLASSICAL ELECTRODYNAMICS
Mathias Frisch

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

The Origin and Evolution of Cultures
ROBERT BOYD and PETER J. RICHERSON
Oxford presents, in one convenient and coherently organized volume, 20 influential but until now relatively inaccessible articles that form the backbone of Boyd and Richerson's path-breaking work on evolution and culture.
January 2005 paper $35 00 cloth $74.00

Inconsistency, Asymmetry, and Non-Locality
A Philosophical Investigation of Classical Electrodynamics
MATHIAS FRISCH
Frisch's overarching philosophical thesis is that standard philosophical accounts that strictly identify scientific theories with a mathematical formalism and a mapping function specifying the theory's ontology are inadequate, since they permit neither inconsistent yet genuinely successful theories, nor thick causal notions to be part of fundamental physics.
January 2005 $49.95

New in Paperback!
Travels in Four Dimensions
The Enigmas of Space and Time
ROBIN LE POIDEVIN
Le Poidevin gives a vivid sense of the difficulties raised by our ordinary ideas about space and time, but he also gives us the basis to think about these problems independently, avoiding large amounts of jargon and technicality. His book is an invitation to think philosophically rather than a sustained argument for particular conclusions.
paper January 2005 $15.95

Speaking My Mind
Expression and Self-Knowledge
DORIT BAR-ON
Dorit Bar-On develops and defends a novel view of avowals and self-knowledge. Drawing on resources from the philosophy of language, the theory of action, epistemology, and the philosophy of mind, she offers original and systematic answers to many long-standing questions concerning our ability to know our own minds.
January 2005 paper $29 95 cloth $74 00

Putting Logic in Its Place
DAVID CHRISTENSEN
Does logic help determine whether beliefs are rational? David Christensen argues that it does—but only once we understand beliefs as coming in degrees. Avoiding mathematical technicality, he explains why the degree-of-belief approach offers the key to understanding how logical arguments work.
January 2005 $39.95

JKC.110104

CAMBRIDGE

OUTSTANDING SCHOLARSHIP FROM CAMBRIDGE

The Creativity of God

World, Eucharist, Reason

Oliver Davies

Theological tradition has lost its sense of the created nature of the world, leading to a complete break between the way we know God and the world. The Johannine tradition of creation through the Word and a Eucharistic semiotic of Christ as the embodied creative speech of God serve to reintegrate Christian cosmology. This book sets out "a new understanding" of reality, as encountered in and through Jesus Christ, and explores new modes of ecclesial reasoning.

Cambridge Studies in Christian Doctrine 12

$75.00: Hardback 0-521-83117-2 222pp.
$27.99: Paperback: 0-521-53845-9

Justice and Democracy

Keith Dowding, Robert E. Goodin, and Carole Pateman, Editors

While much has been written about social justice, even more has been written about democracy. Rarely is the relationship between social justice and democracy carefully considered. Does justice require democracy? Will democracy bring justice? This volume brings together leading authors who consider the relationship of democracy and justice. The intrinsic justness of democracy is challenged and the relationship between justice, democracy and the common good examined.

$70.00 Hardback: 0-521-83695-6. 242pp
$25.99. Paperback. 0-521-54543-9

Myths of the Underworld Journey

Plato, Aristophanes, and the "Orphic" Gold Tablets

Radcliffe G. Edmonds III

Plato, Aristophanes, and the creators of the "Orphic" gold tablets employ the traditional tale of a journey to the realm of the dead to redefine, within the mythic narrative, the boundaries of their societies. Rather than being the relics of a faded ritual tradition or the products of Orphic influence, these myths can only reveal their meanings through this detailed analysis of the specific ways in which each author makes use of the tradition.

$75.00: Hardback: 0-521-83434-1. 288pp.

Alfred Tarski

Life and Logic

Anita Burdman Feferman and Solomon Feferman

Alfred Tarski, one of the greatest logicians of all time, is widely thought of as "the man who defined truth." His mathematical work on the concepts of truth and logical consequence are cornerstones of modern logic, influencing developments in philosophy, linguistics and computer science. This first full-length biography places Tarski in the social, intellectual and historical context of his times. It presents a vivid picture of a personally and professionally passionate man, interlaced with an account of his major scientific achievements.

$35.00: Hardback: 0-521-80240-7 432pp

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 2
FEBRUARY 2005

page
55 *Almost Indiscernible Objects and the Suspect Strategy* Kathrin Koslicki
78 *How to Read "Epistemology Naturalized"* Bredo C. Johnsen
COMMENTS AND CRITICISM
94 *The Meaning of 'Race': Folk Conceptions and the New Biology of Race* Robin O. Andreasen
107 NOTES AND NEWS
108 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc.

ISSN 0022-362X

University Library ALIPORE CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NO. 2, FEBRUARY 2005

ALMOST INDISCERNIBLE OBJECTS AND THE SUSPECT STRATEGY*

In this paper, I argue that a surprisingly widespread strategy in metaphysics is suspect for various reasons and hence ought to be abandoned. In very broad strokes, situations which give rise to *the Suspect Strategy* (TSS) contain as one of their ingredients a general metaphysical principle of some form whose truth the proponent of (TSS) wishes to uphold; the nature of the principle differs from context to context, but examples include the following:

(LL) *Leibniz's Law:*

If $x = y$, then every property of x is a property of y.[1]

(RI) *Restricted Indiscernibility:*

If a certain relation, R, holds between x and y, then every Φ-property of x is a property of y.[2]

---

* I would like to acknowledge the generous support I received from Tufts University and the Andrew Mellon Foundation during the 2003–2004 academic year, which helped to free up time for the completion of this project. A shorter version of this paper was presented at the Spring 2004 Pacific Division meeting of the American Philosophical Association in Pasadena, California; I am grateful to my commentator, Guy Rohrbaugh, as well as the members of the audience for their insightful comments. My work on this paper has also profited from discussions with Jody Azzouni, Lynne Rudder Baker, Harry Deutsch, Graeme Forbes, Michael Glanzberg, Jim Pryor, Mark Richard, Nathan Salmon, Ori Simchen, and Achille Varzi.

[1] For the sake of simplicity, I am omitting relations.

[2] (RI) is a *schema* of which particular restricted indiscernibility principles are instances. As it stands, (RI) contains at least two open places. (i) The place marked by 'R' is to be filled in by a relation which is similar to but weaker than numerical identity (for example, the relation of constitution); if R is taken to be numerical identity, then 'Φ' can be taken to mark no restriction at all, and (RI) simply collapses into (LL). (ii) The family of properties with respect to which the R-related objects are indiscernible must be explicitly specified, that is, 'Φ' must be filled in in some way (for example, in the case of constitution, one will want to exclude the property of being essentially a statue from the family of Φ-properties; such "ordinary" intrinsic

0022-362X/05/0202/55–77

The second ingredient which is needed to give rise to (TSS) is a certain troublesome class of contexts, Σ (for example, contexts like '____ is essentially a statue'). These contexts appear to satisfy the purely formal syntactic and semantic well-formedness conditions expressions must satisfy in order to play the semantic role of predicates. (For example, they are "unsaturated," in Frege's sense: that is, when combined with singular terms, they yield statements that can bear a truth value; they apparently do not lead to paradox, and so forth.) However, to allow that these contexts straightforwardly determine *properties* and that these properties straightforwardly *fall under the scope of* the general metaphysical principle in question would conflict with certain *other* metaphysical priorities of the proponent of (TSS).

To resolve this tension, the philosopher in question invokes (TSS), with the intended result that the troublesome contexts in Σ be *excluded* from the reaches of the general principle in question, either because these contexts fail to determine properties at all or because the properties they do determine fail to fall under the scope of the general principle at issue. What makes the strategy in question *suspect* is that, as we shall see, the different kinds of methods by which the troublesome contexts are excluded from the reaches of the general principles raise serious methodological concerns or are objectionable for other reasons.

We should draw two conclusions from the failure of (TSS). First, the need to invoke (TSS) by itself counts as a strike against a philosophical theory; hence, competing theories which require no such appeal are preferable in this respect. Second, unless other independently motivated considerations are provided, the rejection of (TSS) presents a good reason to accept that the contexts in Σ determine properties and that these properties fall under the scope of the general metaphysical principle in question (provided, of course, that this principle is taken to be true). Though I cannot properly argue for this stronger claim here, this second consequence of the failure of (TSS) in my view further commits us to a universe populated with numerically distinct yet almost indiscernible objects.[3]

I. THE SUSPECT STRATEGY

I now turn to some representative illustrations of contexts in which (TSS) is applied with respect to the general principles mentioned

---

and relational properties as weight and spatiotemporal location, on the other hand, should be included in the family in question).

[3] See also Kit Fine, "The Non-Identity of a Material Thing and Its Matter," *Mind*, CXII (2003): 195–234, for a recent critique of various attempts to block inferences using (LL) which lead to numerically distinct, spatiotemporally coincident objects.

above, (LL) and (RI).[4] For example, we find (TSS) implemented with respect to (LL) in (i) Allan Gibbard's defense of *contingent identity* as well as (ii) Terence Parsons's defense of *indeterminate identity*.[5] An example of (TSS), as implemented with respect to an instance of (RI), occurs in (iii) a recent development of Peter Geach's *relative identity* view by Harry Deutsch.[6]

I.1. The Suspect Strategy and Leibniz's Law. A. Contingent Identity. In his classic paper, "Contingent Identity," Gibbard argues that certain identities are best interpreted as contingent, despite Saul Kripke's powerful arguments to the contrary.[7] As an example of such a contingent identity, Gibbard offers the case of a statue, Goliath, and the piece of clay, Lumpl, of which it is made, which are stipulated to have exactly the same temporal extent; their relation, in Gibbard's view, is best described as in:

---

[4] Moreover, the same strategy as implemented with respect to a different metaphysical principle,

(EP) *Existence Principle:*
For any set of Φ-properties, there exists an object which has all the properties in the set and no other Φ-properties

can be found in Terence Parsons's defense of *nonexistent objects*—"Referring to Non-Existent Objects," *Theory and Decision*, XI (1979): 95–110, and *Non-Existent Objects* (New Haven: Yale, 1980).

[5] Gibbard, "Contingent Identity," *Journal of Philosophical Logic*, IV (1975): 187–221; Parsons, *Indeterminate Identity: Metaphysics and Semantics* (New York: Oxford, 2000). For illuminating discussion of Parsons, see Nathan Salmon, "Identity Facts," *Philosophical Topics*, XXX, 1 (2002): 237–67.

A further implementation of (TSS), as applied to (LL), can be found in the context of George Myro's and André Gallois's defense of *temporary identity*—cf. Myro, "Identity and Time," in Richard E. Grandy and Richard Warner, eds., *The Philosophical Grounds of Rationality* (New York; Oxford, 1986), pp. 383–409; as well as Gallois, "Occasional Identity," *Philosophical Studies*, LVIII (1990): 203–24, and *Occasions of Identity: A Study in the Metaphysics of Persistence, Change, and Sameness* (New York: Oxford, 1998).

[6] Geach, *Reference and Generality* (Ithaca, NY: Cornell, 1962), and "Identity," *Review of Metaphysics*, XXI (1967): 3–12; Deutsch, "Identity and General Similarity," in J.E. Tomberlin, ed., *Philosophical Perspectives*, Volume 12, *Language, Mind and Ontology* (Atascadero, CA: Ridgeview, 1998), pp. 177–99. See also Deutsch, "Relative Identity," in Edward N. Zalta, ed., *Stanford Encyclopedia of Philosophy* (Summer 2002 edition on the internet).

A similar implementation of (TSS) with respect to an instance of (RI) is utilized by some *coincidence theorists*, in their response to the problem of constitution. See, for example, Lynne Rudder Baker, "Unity without Identity: A New Look at Material Constitution," *Midwest Studies in Philosophy*, XXIII (1999): 144–65, and *Persons and Bodies: A Constitution View* (New York: Cambridge, 2000); Fine, "Acts, Events and Things," in Werner Leinfellner, Eric Kraemer, and Jeffrey Schank, eds., *Language and Ontology, Proceedings of the 6th International Wittgenstein Symposium, 1981* (Wien: Hölder-Pichler-Tempsky, 1982), pp. 97–105, and "Things and Their Parts," *Midwest Studies in Philosophy*, XXIII (1999): 61–74; Stephen Yablo, "Identity, Essence and Indiscernibility," this JOURNAL, LXXXIV, 6 (June 1987): 293–314.

[7] Kripke, "Identity and Necessity," in Milton Munitz, ed., *Identity and Individuation* (New York: University Press, 1971), pp. 135–64.

(1) $$\text{Goliath} = \text{Lumpl} \mathbin{\&} \Diamond\,(\text{Goliath} \neq \text{Lumpl})$$

Of course, as Gibbard points out in section v of his paper, one's immediate reaction is that cannot possibly be the correct interpretation of the relation between Lumpl and Goliath, on the grounds of the following style of argument:[8]

(2) $$\begin{array}{c} \Box\,(\text{Lumpl} = \text{Lumpl}) \\ \text{Lumpl} = \text{Goliath} \\ \hdashline \Box\,(\text{Goliath} = \text{Lumpl}) \end{array}$$

The argument in (2) states that because Lumpl is necessarily self-identical, so anything that is identical with Lumpl, namely, Goliath, also must be necessarily identical to Lumpl. This argument depends on taking the context in (3),

(3) $$\Box\,(\underline{\hphantom{xxxx}} = \text{Lumpl})$$

in conjunction with (LL), to generate the conclusion in (2) which contradicts Gibbard's central thesis in (1). (Gibbard in "Contingent Identity" specifically addresses himself to an argument of this sort that is given by Kripke; Kripke uses this argument to conclude that such pairs of objects as Lumpl and Goliath must be numerically distinct.) In other words, if the argument in (2) is correct, then the context in (3) points us to a property with respect to which the objects in question are not indiscernible (namely, necessary identity with Lumpl); (LL) would then seem to lead us to conclude that Lumpl and Goliath are numerically distinct and hence not contingently identical, contra (1).

Gibbard calls this the "most prominent objection" to the contingent identity view; his response is an instance of (TSS):

> The usual answer will serve my purpose here. Leibniz's Law settles very little by itself: put as a general law of substitutivity of identicals, it is just false; in its correct version, it is a law about properties and relations: *If $x = y$, then for any property, if x has it, then y has it, and for any relation and any given things, if x stands in that relation to those things, then y stands in that relation to those things.* The law so stated yields substitutivity of identicals only for those contexts that attribute properties and relations. [The conclusion in (2)] follows from [the two premises] by Leibniz's Law,

---

[8] For simplicity, I omit relativization to existence in this and all following arguments.

> then, only if [the context in (3)] attributes a property. We can block the inference to [the conclusion in (2)] by denying that [the context in (3)] attributes a property.[9]

In case someone should worry about the possible "arbitrariness" of this response, Gibbard remarks that whether the context in (3) denotes a property is precisely what is at issue in the dispute between the essentialist and the anti-essentialist. A context denotes a property, so Gibbard argues (plausibly, of course), only if it applies to an object *independently of the way in which it is designated*; and whether *de re* modal contexts apply to objects in this fashion is precisely the point over which anti-essentialists like Gibbard, W.V. Quine, and others disagree with essentialists like Kripke. The battle between them must therefore be fought on other grounds.[10]

---

[9] Gibbard, "Contingent Identity," p. 201 (his italics; the numbering of examples has been adjusted to my text).

[10] A similar strategy is employed by the temporary identity view, as developed in different ways in Myro and Gallois. The temporal case is slightly more tricky than its modal analogue and depends on using the following instance of $(\mathrm{LL}_{\mathrm{Temp}})$,

(i) $[\text{at } t\text{: } A=B] \rightarrow [[\text{at } t\text{: } \forall t' \ (\text{at } t'\text{: } A=A)] \leftrightarrow [\text{at } t\text{: } \forall t' \ (\text{at } t'\text{: } A=B)]]$

in conjunction with the premise that at a certain time, t, A and B are identical, to infer the consequent of the conditional in (i). Since A is always identical with itself, it is plausible to think that it is also true *at a particular time*, t, that A is always identical with itself; in that case, $(\mathrm{LL}_{\mathrm{Temp}})$, in conjunction with the assumption that the context in (ii) denotes a property,

(ii) At all times t': [at t': $A =$ _____]

permits the inference to the conclusion that if A is *ever* identical to B, then it is so *always*, which contradicts the temporary identity view.

Myro and Gallois respond to the challenge posed by the argument from $(\mathrm{LL}_{\mathrm{Temp}})$ by endorsing slightly different versions of (TSS). Myro's response is in fact quite close to Gibbard's:

> So the general way of dealing with the complication is to divide properties into those which are *'time-free'*—like being on the mantelpiece—which are represented by open sentences *not* containing temporal qualifications, and those which are *'time-bound'*—like being on the mantelpiece on Tuesday—which are represented by open sentences which do contain temporal qualifications. And what must be done is that 'Leibniz's Law subject (like other statements) to temporal qualification' is to be, in addition, *restricted* to properties which are *'time-free'*—properly represented by open sentences (or 'predicates') which do not (relevantly) contain temporal qualifications ("Identity and Time," pp. 392–93, italics added).

Gallois blocks the inference in question, not by overtly restricting $(\mathrm{LL}_{\mathrm{Temp}})$ or by openly declaring that contexts of a certain kind fail to denote properties, but rather by opposing a certain pretheoretically plausible principle concerning the transmission of properties through times:

(TP) $(\forall F)\ (\forall x)\ (\forall t)\ (\forall t')\ [\text{at } t'\text{: } Ex \rightarrow [\text{at } t\text{: } F(x) \leftrightarrow [\text{at } t'\text{: } F(x) \text{ at } t]]]$

By rejecting (TP), the following instance of this principle, which was appealed to above,

(iii) $\forall t\ (\text{at } t\text{: } A=A) \rightarrow \forall t'\ [\text{at } t'\text{: } \forall t\ [\text{at } t\text{: } A=A]]$

I.1.B. Indeterminate Identity. Parsons in his *Indeterminate Identity* defends the view that, under certain circumstances, identities can be indeterminate; that is, that statements of the following kind can be true (where the operator, '∇' is taken to mean 'it is indeterminate that'):

(4) $\nabla\ (B = A)$

Once again, the defender of indeterminate identities faces an objection from (LL), structurally analogous to that reviewed above, except for the fact that the argument in question this time makes use of (LL) in its *contrapositive* form (see Gareth Evans for the original statement of this argument):[11]

($\mathrm{LL}_{\mathrm{Contra}}$) *Contrapositive Leibniz's Law:*
If some property, F, is a property of x but not of y, then $x \neq y$.

The identity sign, '=', is read by the defender of indeterminate identity as applying to objects which are *determinately* identical; correspondingly, '$\neq$' applies to objects which are *determinately* distinct. Normally, the equivalence between (LL) and ($\mathrm{LL}_{\mathrm{Contra}}$) is of course taken for granted. In the context of disputes over the determinacy of identity, however, this equivalence is no longer uncontroversial; Parsons, for example, accepts (LL) but denies that inferences using ($\mathrm{LL}_{\mathrm{Contra}}$) are always valid.

Now assume, for reductio, that objects, A and B, are indeterminately identical, that is, that (4) is true. Then, the argument from ($\mathrm{LL}_{\mathrm{Contra}}$) can be stated as follows:

(5)
$$\begin{array}{c} \nabla\ [B = A] \\ \neg\nabla\ [A = A] \\ \hline B \neq A \end{array}$$

The argument in (5), again, proceeds by way of taking contexts like (6),

(6) $\nabla\ [\_\_\_\_ = A]$

---

is now no longer available. Although Gallois's careful treatment of the issues in question deserves separate discussion, his position nevertheless strikes me in the end as a slightly more elaborate version of the view that there is no automatic passage from contexts of a certain purportedly questionable kind to properties of the corresponding kind, where the contexts in question are now those involving *nested* temporal qualifications; for insightful and detailed discussion of Gallois' views, see Theodore Sider, *Four-Dimensionalism: An Ontology of Persistence and Time* (New York: Oxford, 2001).

[11] Evans, "Can There Be Vague Objects?," *Analysis*, XXXVIII (1978): 208.

in conjunction with (LL_{Contra}), to lead to the conclusion in (5), according to which *A* and *B* are determinately distinct, which contradicts the assumption in (4). This argument is used by the opponent of indeterminate identity to show that objects can never be merely indeterminately identical; that is, that identity is always determinate.

In response to this Evans-style argument, Parsons proposes the familiar strategy of denying that contexts like that in (6) denote properties. He does, however, introduce a novel consideration in support of his version of (TSS). What makes contexts like (6) suspicious, according to Parsons, is that they bear some structural similarity to contexts which are used to generate the paradoxes of naive set theory. Since Parsons also accepts that (determinate) identity can be *defined* as the sharing of properties as in (7),

(7) $$A = B \equiv_{\text{def}} \forall P\, [P(A) \leftrightarrow P(B)]$$

contexts like (6), in his view, involve implicit quantification over all properties. Parsons explains the analogy between the Evans-style argument and set-theoretic paradoxes as follows:

> The force behind the reasoning thus comes from the fact that identity is defined in terms of what properties there are, and a problematic property is defined using an abstract that quantifies over *those* properties. The condition in the abstract is cleverly designed to conflict with its yielding one of the properties quantified over (if any objects are indeterminately identical with A). The reasoning thus resembles that of the Russell paradox in set theory. (Identity between sets is defined in terms of what sets they have as members, and a problematic set is defined using a set abstract that quantifies over *those* sets. The condition in the set abstract is cleverly designed to conflict with its yielding one of the sets quantified over.)[12]

Given the analogy with the paradoxes of naive set theory, Parsons takes himself to be justified in adopting his version of (TSS), namely, that contexts which have this apparently impredicative character cannot always be expected to determine a property.

I.2. The Suspect Strategy and Restricted Indiscernibility Principles: Relative Identity. The third context I want to consider occurs in a subtle and interesting recent development of Geach's relative identity view in Deutsch. According to Deutsch, the relative identity theory can solve many classical metaphysical problems that concern numerical identity in an attractive way; examples he considers include the following:

---

[12] Parsons, *Indeterminate Identity*, p. 51.

(8) *Metaphysical Puzzle Cases:*
Change over Time: 'The young Fido is *the same dog as* the old Fido'.
Constitution: 'Lumpl is *the same statue as* Goliath'.
Types and Tokens: 'My copy of *On the Road* is *the same literary work* as that originally written by Kerouac'.

In each case, Deutsch proposes that the relation in question, for example, *being the same dog as, being the same statue as,* and *being the same literary work as,* is best analyzed as a relation of *relative identity.* Thus, the relation in question does not dissolve, as the absolute identity theorist would have it, into a predicative component and a component that denotes absolute identity, as in 'x is a dog and y is a dog and $x \doteq y$'; rather, the relation in question is not further analyzable and simply denotes a feature of the world, namely, one of the ways in which objects that are numerically distinct in the absolute sense can be similar to one another. (Unlike Geach, Deutsch does not believe that absolute identity is incoherent or unintelligible and accepts that objects which are merely relatively identical are numerically distinct in the absolute sense.)

As Deutsch acknowledges, any plausible version of the relative identity theory must respond in some manner to David Wiggins's original challenge to Geach: to offer a suitable *restricted* indiscernibility principle which can be said to govern relative identity in place of the unrestricted version of (LL).[13] For if Lumpl and Goliath are not the same statue in the *absolute* sense, we of course have no right to expect them to be indiscernible in absolutely *every* respect, as (LL) would have it. But we do have a right to ask how the relative identity theorist will explain the fact that being similar in *this* respect (namely, the respect denoted by 'is the same statue as') entails being similar in so many *other* respects, in an entirely predictable and systematic fashion: statues and the objects that constitute them can *always* be expected to have the same weight, shape, color, texture, chemical composition, and so forth. Thus, the relative identity theorist bears the responsibility of offering a restricted indiscernibility principle of some kind, as in (RI_{Rel}),

(RI_{Rel}) *Restricted Indiscernibility of Relatively Identical Objects:*

If x is relatively identical to y, then every Φ-property of x is a property of y.

which will, among other things, validate inferences like those in (9),

---

[13] David Wiggins, *Sameness and Substance* (Cambridge: Harvard, 1980), pp. 18ff., and *Sameness and Substance Renewed* (New York: Cambridge, 2001), pp. 24ff.

(9) Lumpl has the Φ-property *F*.
Lumpl is the same statue as Goliath.

Goliath has *F*.

The crucial question for the relative identity theorist is how to fill in 'Φ' in such a way as to *exclude* troublesome contexts such as the following,

(10) *Troublesome Contexts:*
Modal: '_____ is essentially a piece of clay'
Temporal: '_____ existed before the statue came into existence'
Identity: '_____ is (absolutely) identical to the lump of clay'
Constitution: '_____ constitutes a statue'

be *excluded* from the reaches of the restricted indiscernibility principle in ($\mathrm{RI}_{\mathrm{Rel}}$), since they will in general *invalidate* inferences like those in (9). Only the task faced by the relative identity theorist is especially challenging, since 'Φ' must be filled in in such a way that it will simultaneously validate inferences in *all* the metaphysical contexts for which relative identity is intended to yield an analysis, for example, contexts involving the phenomenon of *change over time* as well as those involving *constitution* and the *identity of allographic objects.*

As his version of ($\mathrm{RI}_{\mathrm{Rel}}$), Deutsch proposes the principle he calls '(T4)' which is here reworded in a more informal fashion (for reasons which shall become apparent momentarily, I label this principle, Deutsch's *Expansion Principle*):

($\mathrm{RI}_{\mathrm{Rel\text{-}Deutsch}}$) *Deutsch's Expansion Principle:*

> If *x* is the same *F* as *y*, then *y* has all of those properties of *x* which satisfy the condition: if *some F* has the property in question, then *all* the *F*'s do.

The intuitive idea behind ($\mathrm{RI}_{\mathrm{Rel\text{-}Deutsch}}$) is to isolate those properties which "spread through" the entire equivalence class singled out by a particular relative identity relation. For example, consider the equivalence class consisting of all the different objects (numerically distinct, in the absolute sense) which are *the same statue as* Goliath (at a particular time or over time). The Φ-properties with respect to this equivalence class are those which satisfy the condition: if one such "Goliath object" has the property in question, then they all do. As we shall see below, Deutsch's version of ($\mathrm{RI}_{\mathrm{Rel}}$) compares favorably, from a methodological point of view, to other strategies of excluding the troublesome contexts; but it is nevertheless suspect for other reasons.[14]

---

[14] A similar implementation of (TSS) with respect to an instance of (RI) can be found in the coincidence theorist's response to the problem of constitution, as developed for example in Baker, Fine, and Yablo. To explain the striking similarities

II. WHAT IS WRONG WITH THE SUSPECT STRATEGY?

In our illustrations of (TSS) above, we have encountered basically four different strategies of how to exclude the troublesome contexts from the reaches of the general principle at issue. (i) First, there is what I shall term the *Purely Stipulative Strategy*; this strategy is the most widespread in the literature and is here exemplified by Gibbard.[15] (ii)

---

between constitutionally related objects, Baker, Fine, and Yablo each propose slightly different versions of a restricted indiscernibility principle of the following form:

(RI_{Const}) *Restricted Indiscernibility of Constitutionally Related Objects:*
If x constitutes y, then every Φ-property of x is a property of y.

The differences between the three accounts lie in precisely how 'Φ' is to be filled in; I will return to this question in more detail below (for further discussion, see also my "Constitution and Similarity," *Philosophical Studies*, CXVII (2004): 327–64). Like the relative identity theorist, the coincidence theorist must explain the validity of inferences that are analogous to those considered earlier in the context of our discussion of (LL),

$$\begin{array}{l} \text{Lumpl has the } \Phi\text{-property } F. \\ \text{Lumpl constitutes Goliath.} \\ \hline \text{Goliath has } F. \end{array}$$

while *excluding* from the reaches of the restricted indiscernibility principle in (RI_{Const}) troublesome contexts such as those mentioned in section I.2, since they will in general *invalidate* these inferences. As we will see below, the strategy used by Baker, Fine, and Yablo to exclude the troublesome contexts in question from the general metaphysical principle in (RI_{Const}) is structurally analogous to earlier implementations of (TSS), especially those encountered in Gibbard and Myro.

In Terence Parsons's defense of nonexistence objects ("Referring to Non-Existent Objects" and *Non-Existent Objects*), we see an application of (TSS) with respect to the existence principle, (EP), mentioned earlier (see note 4): according to this principle, a certain special class of properties is such that, for any set of them, there is an object that (determinately) has all of the properties in question (and is indeterminate with respect to all other properties in the class). The special, or "nuclear," properties in question, in Parsons's view, are those denoted by such predicates as 'is blue', 'is tall', and 'was kicked by Socrates'; they contrast with those that are "extra-nuclear," for example, those denoted by such predicates as 'exists', 'is mythical', 'is fictional', 'is possible', 'is impossible', 'is thought about by Meinong', and 'is worshipped by someone'. In response to the crucial question of how this list is to be continued, Parsons remarks:

> Our historical situation yields a very rough kind of decision procedure for telling whether a predicate is nuclear or extranuclear. It's this: if everyone agrees that the predicate stands for an ordinary property of individuals, then it's a nuclear predicate, and it stands for a nuclear property. On the other hand, if everyone agrees that it doesn't stand for an ordinary property of individuals (for whatever reason), or if there's a history of controversy about whether it stands for a property of individuals, then it's an extranuclear predicate, and it does not stand for a nuclear property ("Referring to Non-Existent Objects," p. 102).

As in earlier implementations of (TSS), Parsons proceeds by simply legislating that the extra-nuclear properties (whatever they are) be excluded from the reaches of his general principle, (EP).

[15] But see also Myro; Gallois; the Parsons of nonexistent objects; Baker; Fine; and Yablo.

Second, we see in Gibbard an extremely condensed further suggestion which, if it were elaborated more fully, might seem to point the way towards a nonstipulative response; I shall term this Gibbard's *Appeal to Failures of Substititutivity.* (iii) Third, we came across a novel and intriguing suggestion in the Parsons of indeterminate identity, namely, that the troublesome contexts in question are somehow analogous to those that give rise to the paradoxes of naive set theory and should be excluded from the reaches of the general principle on those grounds; I shall term this response Parsons's *Appeal to the Paradoxes of Naive Set Theory.* (iv) Finally, we considered a creative proposal by Deutsch on behalf of the relative identity theorist, which I have labeled Deutsch's *Expansion Principle.* In what follows, it will be my aim to show that none of these strategies of excluding troublesome contexts from the reaches of the general principle is successful.

II.1. The Purely Stipulative Response. I turn, first, to the Purely Stipulative Strategy, which is to be found for example in Gibbard. The Purely Stipulative Strategy proceeds by way of excluding, on *purely stipulative grounds,* a set of troublesome contexts from the reaches of a general metaphysical principle whose truth the philosopher in question wishes to uphold: it is simply legislated either that these contexts *fail to denote properties* altogether or that the properties they do denote *fail to fall under the scope* of the general metaphysical principle in question; Gibbard takes the first, less common, strategy.

Of course, the mere fact that this strategy is purely stipulative makes it seem ad hoc and hence methodologically suspect. I will, however, try to say more explicitly what it is about this strategy that should worry us, since its proponents might suggest that *some* purely stipulative maneuvers are worth their philosophical price. What makes the Purely Stipulative Strategy especially troubling is that it has the following feature. In each case, there is only a handful of contexts which, when combined with the general metaphysical principle at issue, will generate trouble for the philosopher in question. For example, in the case of the contingent identity theorist, the general principle is (LL) in its unrestricted, nontemporalized form, and the contexts in question are only those that would conflict with the thesis that coincident objects with the same spatiotemporal extent are contingently identical, for example, contexts of the following sort (or whatever else the essentialist wishes to substitute):

(11) *Troublesome De Re Modal Contexts:*
Necessary Identity: ☐ (_____ = *A*)
Essential Kind Membership: ☐ (_____ is a statue).
Essentiality of Origin: ☐ (_____ was fashioned by artist so-and-so)

In response to the potential threat posed by contexts like those in (11), Gibbard adopts the view that *de re* modal contexts *in general* fail to denote properties. This strategy has momentous consequences, as it leads to a complete reinterpretation of much of our discourse: it requires, among other things, a new theory of proper names, a new notion of rigidity, a new conception of crossworld identity and a new conception of what goes on in contexts in which we seem to attribute *de re* modal properties to concrete objects directly. It does, however, achieve the intended result of effectively removing the troublesome contexts from the reaches of (LL), since, as Gibbard remarks, (LL) is to be understood as a *metaphysical* principle ranging over objects, properties, and relations, and not as a *linguistic* principle of substitutivity ranging over contexts and expressions.

The difficulty for the contingent identity theorist now is that there are plenty of contexts which satisfy the purely formal criteria of being *de re* modal (namely, they involve an occurrence of a name or unbound variable within the scope of a modal operator), and which are completely harmless from the point of view of the contingent identity theorist, in the following sense: if they *were* to be included in the scope of (LL), they would *not* conflict with the thesis of contingent identity; I have in mind contexts of the following sort (assuming, with Gibbard, that dispositional, counterfactual, and causal contexts involve *de re* modality):

(12) *Harmless De Re Modal Contexts:*
Dispositional: (____ is fragile)
(____ conducts electricity thus-and-so)
Counterfactual: (if ____ were dropped on my foot, my foot would swell)
Causal: (____ prevents my hand from passing through it)
(____ casts a shadow of length so-and-so when hit by the sun at angle thus-and-so)

If the contingent identity theorist were to exclude from the reaches of (LL) *only* the contexts in (11), and *not* those in (12), then the arbitrariness of his strategy would presumably be just too blatant: contexts would then be sorted into those which fall under the scope of (LL) and those which fail to do so simply by whether the result would conflict with the contingent identity theory.

To avoid this undisputably blatant arbitrariness, Gibbard adopts a more coarse-grained individuation criterion for troublesome contexts, which includes *all* contexts that satisfy the purely formal criteria for being *de re* modal; that is, the harmless contexts in (12) along with the troublesome contexts in (11). In his very condensed remarks in section v of his paper (some of which were quoted above), Gibbard seems to suggest that this more coarse-grained individuation criterion can actually be justified on independent grounds, namely, on the grounds that *de re* modal contexts in the eyes of the anti-essentialist fail to satisfy a generally plausible principle governing the relation between linguistic contexts and properties:

(13) *Independently Plausible Principle concerning Property Formation:*

A context denotes a property only if it applies to an object *independently of how the object is designated.*

I will comment in more detail below on why I do not believe that (13) succeeds in accomplishing its intended goal. For now, I want only to note that the exclusion procedure Gibbard adopts in the interest of avoiding the undisputably blatant arbitrariness yields the wrong results by virtue of being *too* coarse-grained. For by excluding the harmless contexts in (12) from the reaches of (LL), along with the troublesome contexts in (11), the contingent identity theorist has now done away with contexts with respect to which contingently identical objects can in general be *expected* to be indiscernible. If (LL) can no longer be used to provide an explanation of this datum, then some *other* explanation must take its place. This, of course, puts the contingent identity theorist in exactly the same boat as the coincidence theorist (see note 14) and the relative identity theorist: for he is now in need of a *restricted* indiscernibility principle like (RI) (only one that is formulated in terms of contexts rather than properties), which provides a systematic account of the ways in which contingently identical objects are indiscernible. This principle, again, must be formulated in such a way as to *exclude* the troublesome contexts in (11) and *include* the harmless contexts in (12).

But how do we formulate such a principle in a way that is not methodologically or otherwise suspect? Gibbard cannot help himself to the strategy adopted by those coincidence theorists, such as Lynne Rudder Baker, Kit Fine, and Stephen Yablo, who invoke a restricted indiscernibility principle to explain the striking similarities between constitutionally related objects; for their proposals suffer from exactly the same weakness of Gibbard's own.[16] (Deutsch's idea of how to proceed will be considered separately below.) For, according to Fine,[17] the family of properties over which the relevant indiscernibility principle extends is defined to include all and only those that are "normal," where a "normal" property is one that is not "formal" and whose application concerns only the *time* and *world* in question. The notion of a "formal" property is not further elucidated by Fine, but I take it to include such purely "logical" properties as the property of being self-identical and the property of being either red or not red.[18] Baker defines the family of properties in question in a similar fashion, as those that include all properties *except* those that are ("alethic") *modal*, those that concern *identity* and *constitution*, and those that are "rooted outside" the times at which they are had (*op. cit.*). For Yablo, they are all and only those properties that are "categorical," that is, roughly those that concern what goes on in the *actual* world; the properties that are *excluded* from the family in question are the "hypothetical" ones, that is, those that concern what goes on in *other* worlds.

---

[16] As discussed in my "Constitution and Similarity," Fine, Baker, and Yablo are actually quite unusual among coincidence theorists, in that they pay any attention at all to the problem of how to capture the striking similarities among constitutionally related objects.

But these proposals suffer from the same weaknesses as Gibbard's own, in that they are (i) *purely stipulative* and (ii) *overly coarse-grained.* They are purely stipulative, because it is simply *legislated* that contexts of the troublesome kind are to be excluded from the reaches of (RI), without any attempt at giving an independent justification for why *these* properties, and not others, deserve this special status with respect to the principle at issue. Moreover, by using purely formal criteria of individuating contexts (for example, the occurrence of particular operators in certain syntactically defined ways), these proposals draw the boundaries in the wrong place: they fail to distinguish between the harmless contexts in (12) and the troublesome contexts in (11), since both involve *de re* modal attributions. Thus, unless less coarse-grained methods of delineating contexts can be found, we should be skeptical that the Purely Stipulative Strategy can be made to work.

II.2. Gibbard's Appeal to Failures of Substitutivity. With his very condensed remarks in section v of his paper, Gibbard suggests that the anti-essentialist in fact has independent motivation for removing the troublesome contexts from the reaches of (LL), by virtue of the general principle in (13) cited above which is to govern the relation between linguistic contexts and properties. It is not entirely clear how Gibbard imagines that (13) will help the contingent identity theorist with respect to the "most prominent objection" coming from (LL); in what follows, I lay out what I take to be his implicit reasoning.

---

[17] See "Acts, Events, and Things."

[18] A similar principle is also to be found as "Postulate (V7)" in Fine's "Things and Their Parts."

In addition to the *metaphysical* principle, (LL), governing objects, properties, and relations, there is also a *linguistic* principle concerning the substitutivity of coreferential expressions, which is sometimes called by the same name and occasionally even taken to be the same principle as (LL); I shall call this principle the *Substitutivity of Coreferring Expressions* (SCE):

(SCE) *The Substitutivity of Coreferring Expressions:*

> For all expressions, α and β, ⌜α = β⌝ expresses a true proposition only if substitution of α for β is truth-preserving.

The phrase, 'substitution of α for β is truth-preserving', in (SCE) is to be understood as expressing the following condition:

(TPS) *Truth-Preserving Substitution:*

> For all expressions, α and β, substitution of α for β is truth-preserving if and only if, for all sentences, *S* and *S*′, if *S*′ is like *S* save for containing an occurrence of β where *S* contains an occurrence of α, then *S* expresses a true proposition only if *S*′ does also.[19]

Gibbard remarks that the linguistic principle in (SCE), as it stands, is simply false, and we can concur with him in his assessment, as the evidence to this effect is quite massive and convincing. Counterexamples to (SCE) are drawn primarily from contexts which are considered to be "opaque" in some fashion, for example, 'so-called' constructions such as the following:

(14) Giorgione is so-called because of his size.

(15) Barbarelli is so-called because of his size.

However, none of the counterexamples to (SCE), as Gibbard correctly notes, are thought to affect the truth of (LL): when properly understood, the sorts of considerations that are appealed to in order to reveal the falsity of (SCE) do not present us with cases in which one and the same object is said both to possess and not to possess a single property. For example, the truth of (14) and the falsity of (15), can hardly be used to conclude that the context '____ is so-called because of his size' determines a single property, which one and the same object (that is, the object variously referred to as either 'Giorgione' or 'Barbarelli') both has and lacks. In fact, (LL) is taken by many to be a principle, much like the Principle of Noncontradiction, whose truth is so obvious and fundamental that nothing of an informative and nonquestion-begging nature could be said to justify it. Anything that, on the face of it, looks like a counterexample to (LL) must thus simply involve some sort of misunderstanding.[20]

---

[19] These formulations are taken from Richard Cartwright, "Identity and Substitutivity," in Milton Munitz, ed., *Identity and Individuation* (New York: University Press, 1971), pp. 119–33; reprinted in *Philosophical Essays* (Cambridge: MIT, 1986), pp. 135–47; this citation from p. 136 of the reprinted version.

If my interpretation of Gibbard's reasoning in section v of his paper is correct, then his thought is that, for the anti-essentialist, troublesome contexts like (3),

(3) $\Box(\underline{\quad\quad} = \text{Lumpl})$

are, in the relevant respects, just like "so-called" contexts, in that both involve hidden reference to linguistic expressions. For to be so-called because of one's size is to be called by some *name* or other because of one's size. Similarly, for the anti-essentialist of Gibbard's stripe, an occurrence of a name within the scope of a modal operator as in (3) induces a *nonstandard* interpretation of the name, according to which it is taken to refer to a concrete object not directly, but only *via a sortal concept* of some sort, in this case something along the lines of 'lump of clay'. For objects in and of themselves, according to the anti-essentialist, do not have particular features necessarily or contingently; they do so only *as designated in a certain way.*

On this conception, then, a context like (3) may both apply and fail to apply to one and the same object, depending on whether the single object in question is designated under the name 'Lumpl' or under the name 'Goliath'. And this feature is of course precisely the mark of a context which, according to the independently plausible principle (13), *fails* to determine a property. In this way, so the anti-essentialist reasons, contexts like (3) can at most be used to provide yet another counterexample to the already disproven linguistic principle in (SCE), but they have no relevance to metaphysical principle in (LL).

With Gibbard's reasoning reconstructed in this way, we can now see why the Appeal to Substitutivity does not provide independent motivation for (TSS). My argument comes from three essays by Richard Cartwright in which he demonstrates that the falsity of the linguistic principle in (SCE) has in fact no bearing on the debate between the essentialist and the anti-essentialist.[21] Cartwright's argument, very briefly, is as follows.

---

[20] For arguments to this effect, see, for example, Cartwright, and Mark Richard, "Quantification and Leibniz's Law," *Philosophical Review*, XCIV, 4 (1987): 555–78.

There is actually an important *disanalogy* between contexts like those in (3) and contexts like those in (14) and (15), which we can all agree provide a counterexample to the linguistic principle in (SCE). For suppose we succeed in identifying a 'so-called' context which is in fact both true and false of a single object, depending on whether the object is designated as 'Giorgione' or as 'Barbarelli'; suppose further the context in question is '____ is so-called because of ____'s size'. Then, on pain of *incoherence*, the context in question cannot be said to determine a property, since, in addition to the places marked by '____', it contains another empty place marked by 'so' which has yet to be filled in. Thus, there is *no one* property determined by the context '____ is so-called because of ____'s size'; rather, there are *lots* of properties, depending on how the place marked by 'so' is filled in, which have been misleadingly collected under the same heading: there is the property an object has if it is called 'Giorgione' because of its size; the property an object has if it is called 'Barbarelli' because of its size; and so on. However, once the hidden place marked by 'so' has been explicitly filled in, so that we have in fact succeeded in determining a property, we are no longer dealing with a context which both applies and fails to apply to a single object; depending on how the object is designated. For '____ is called 'Giorgione' because of ____'s size' truly applies to the object in question, no matter how it is designated; and '____ is called 'Barbarelli' because of ____'s size' fails to apply to the object in question, no matter how it is designated. This is the reason why 'so-called' constructions only provide a counterexample to (SCE) but not to (LL).

In a similar vein, the anti-essentialist (according to the version of this view currently under consideration) conceives of *de re* modal contexts like (3) as containing a hidden ellipsis which must be filled in, in this case, by a particular sortal concept before the context in question succeeds in determining a property. For example, the context '____ is necessarily identical to Lumpl', on this view, again denotes a multiplicity of properties, as in '____, *when designated as a lump of clay*, is necessarily identical to Lumpl', '____, *when designated as a statue,* is necessarily identical to Lumpl', and so on. Once a context has been filled in in this way, we will again no longer be faced with a property which both applies and fails to apply to a single object; for it is true of the single statue-shaped object in the actual world, independently of whether it is designated as 'Lumpl' or as 'Goliath', that, *when designated as a lump of clay,* it is necessarily identical to Lumpl; and it is false of the single statue-shaped object in the actual world that, *when designated as a statue,* it is necessarily identical to Lumpl. In this way, the anti-essentialist avoids any conflict with the metaphysical principle, (LL).

---

[21] Cartwright, "Some Remarks on Essentialism," this JOURNAL, LXV, 20 (October 24, 1968): 615–26, "Indiscernibility Principles," *Midwest Studies in Philosophy*, IV (1979): 293–306, and "Identity and Substitutivity."

The essentialist, on the other hand, takes a different view of modal contexts like those in (3). For him, such contexts contain no hidden ellipsis: thus, a context like '____ is necessarily identical to Lumpl', all by itself, that is, without the help of any sortal concept, already succeeds in specifying a property which either applies or fails to apply to an object. And, since Lumpl and Goliath are numerically distinct objects, according to the kind of philosopher we are imagining, there is again no conflict with (LL), since the property determined by '____ is necessarily identical to Lumpl' does not truly apply and fail to apply to a single object.

What makes the situation with respect to such modal contexts as (3) different from that of the agreed-upon counterexamples to (SCE), however, is that, on pain of begging the question against their opponent, neither the anti-essentialist nor the essentialist can appeal to any sort of *incoherence* in the other's position. For the core of the disagreement between them lies precisely in whether *de re* modal contexts like (3) apply to objects in and of themselves, independently of how they are designated. To show that one of the two sides in this dispute is to be preferred over the other, one must appeal, as Gibbard in fact does, to independent, substantive, considerations, for example, the thesis that the essentialist is committed to an unattractive "ghostly" conception of physical objects or that he relies too heavily on questionable modal intuitions. The falsity of the linguistic principle in (SCE) and the plausibility of the principle concerning property formation in (13), however, can do nothing to resolve the dispute between the essentialist and the anti-essentialist; for the two parties can perfectly well agree on all of the following points: (i) that the linguistic principle in (SCE) is false; (ii) that (SCE) is shown to be false, among other things, by contexts like the 'so-called' constructions; (iii) that none of this affects the truth of (LL); and (iv) that the principle in (13) states a correct constraint on property formation. What they disagree on is whether (13) is applicable to *de re* modal contexts like (3); but this disagreement is independent of (i)–(iv). In short, whatever the plausibility of Gibbard's *other* considerations in favor of the contingent identity theory, the falsity of the Substitutivity of Coreferring Expressions principle is simply irrelevant to the dispute between the essentialist and the anti-essentialist.[22]

II.3. Parsons's Appeal to the Paradoxes of Naive Set Theory. Parsons's Appeal to the Paradoxes of Naive Set Theory has the advantage of being methodologically more satisfying than the Purely Stipulative Strategy, since it introduces a systematic, independently motivated consideration by which contexts are to be classified: their apparently vicious impredicative character. It is, however, questionable that the contexts in question really are analogous to those that generate the paradoxes of naive set theory. For note, first, that Parsons's suggestion depends crucially on the assumption that identity can be *defined* as indiscernibility in all respects; unless we accept that the questionable contexts in fact do involve quantification over all properties, they would not be of the allegedly problematic form in which an entity is introduced by means of a definition that quantifies over a domain of elements which is already supposed to include the entity to be defined. By most philosophers' lights, a second order principle in the manner of (7) is unproblematic only if numerical identity is itself included among the properties to be quantified over; if numerical identity is not so included, then the truth of the principle depends on the very controversial assumption that there can be no numerically distinct, qualitatively indiscernible objects. It is therefore open to the opponent of indeterminate identity to block Parsons's reasoning at this point by resisting the *definition* of identity as indiscernibility in all respects.

---

[22] Cartwright's "Indiscernibility Principles" contains a further, powerful objection against Gibbard's particular style of anti-essentialism. Cartwright argues in this essay that the question of whether a context denotes a property is entirely *irrelevant* to the question of whether the corresponding indiscernibility principle is true; for, according to Cartwright, *all* (coherently formulated) indiscernibility principles are true, independently of whether the contexts that occur in them denote properties, and the principle we are accustomed to single out under the name "Leibniz's Law" has no special status among these indiscernibility principles. Gibbard may of course respond to this objection by adopting the more common position of conceding that the contexts in question denote properties, while nevertheless insisting on their exclusion from (LL): however, this concession would not only force a drastic reorientation in many of his other commitments; Gibbard would then still be faced with the task of having to explain why this exclusion of properties from (LL) ought not to be viewed as suspect.

One may worry that my reconstruction of Gibbard's condensed reference to (SCE) results in a position that is not the most favorable to the anti-essentialist; perhaps, the anti-essentialist is better off adopting a position that relativizes *de re* modal contexts in a less overtly *linguistic* manner. In that case, however, one wonders why it is pertinent at all, in this otherwise thoroughly metaphysical context, to point to the falsity of the linguistic substitution-principle as well as the independently plausible constraint concerning property formation.

Moreover, Parsons's analogy is also questionable in the following further respect. Suppose we were to accept that inferences using (LL_{Contra}) are valid, that contexts like (6) denote properties and that identity can be defined in terms of quantification over all properties. Then, the only thing that follows from these assumptions is the conclusion of the Evans-style argument against the possibility of indeterminately identical objects; since the object, *A*, determinately shares all properties with itself, any object which does not determinately share all properties with *A* must be *determinately distinct* from *A*. But no *paradox* ensues from jointly accepting these assumptions. Thus, it seems that Parsons's strategy suffers from the same weakness as Gibbard's Appeal to Failures of Substitutivity, in that it introduces a consideration that is simply irrelevant to the purpose at hand.

Finally, Parsons's strategy, like the Purely Stipulative Strategy above, unsurprisingly also suffers from the weakness of being *overly coarse-grained*, since it as well uses purely formal criteria of individuation (namely, the occurrence of a universal quantifier ranging over properties among which the property to be defined is itself included). Even if we were to grant that *some* contexts involving impredicative definitions lead to paradox, it seems that there are again plenty of *other*, completely harmless, contexts which share the same formal characteristics. For example, suppose an object, *A*, and an object, *B*, have exactly the same number of properties; then, presumably, the context '—— has the same number of properties as *A*' specifies a property which is itself included among *B*'s properties, and correspondingly for *A*. But there is nothing paradoxical about this sort of property.[23]

II.4. Deutsch's Expansion Principle. The final proposal I want to consider is Deutsch's restricted indiscernibility principle governing objects that are identical merely in the relative sense. Such objects, as we know from ($RI_{Rel-Deutsch}$), must share all those properties which, if instantiated by *any* members of a particular equivalence class, must be instantiated by *all* the members of this class.

---

[23] A similar lesson may be learned from those solutions to the semantic paradoxes which trace their source to the phenomenon of self-reference and proceed by legislating that all such contexts are disallowed. Since not all self-reference is problematic (for example, 'This sentence is true'), an approach which proceeds by way of such purely formal criteria of individuation contexts tends to rule out too much.

Like the Parsons of indeterminate identity, Deutsch's proposal is methodologically less suspect than the Purely Stipulative Strategy, in that it introduces a completely general, systematic constraint on (LL); it does, however, suffer from the other weakness we have identified, namely, that of being *overly coarse-grained.* To see why, consider the equivalence class containing all those objects (numerically distinct, in the absolute sense) that are the same literary work as Jack Kerouac's *On the Road* (at a particular time or over time). This equivalence class will consist of a highly nonuniform collection of objects: yellowed paperback copies with missing pages that smell of cigarette smoke and have torn covers, coffee stains, and scribbles in the margins; pristine and beautifully illustrated hardcover, first edition collectors' items, signed by the author; and so on. The regions of space-time occupied by the books themselves are also of course inhabited by the various quantities of matter that constitute them: quantities of paper, cardboard, printer's ink, glue, fabric, and the like. Since Deutsch invokes the relative identity theory to solve the problem of the identity of allographic objects as well as the problem of change over time and the problem of constitution, the different copies of the book themselves as well as the quantities of matter coincident with them are all assigned to the same equivalence class, namely, the class unified by the *being-the-same-literary-work-as* relation. If we now apply Deutsch's Expansion Principle, $(\mathrm{RI}_{\text{Rel-Deutsch}})$, to this heterogeneous bunch, we find that the only properties that satisfy it are properties of a rather *general* sort, namely, those that are commonly taken to be *essential* properties of the literary work in question: for example, kind properties, such as '____ is a book', '____ is an artwork', '____ is an artifact'; origin properties, such as '____ was authored by Jack Kerouac'; and the like. And while Deutsch's principle perhaps says as much as any principle *of logic can* say about the ways in which relatively identical objects can generally be expected to be indiscernible, it would not, for example, satisfy the philosopher who was looking for a response to Wiggins's challenge. For such a philosopher wants to know, for example, when, in general, inferences like those in (9), can be expected to be valid; but Deutsch's principle does not tell us why constitutionally related objects in particular *always* share the same weight, shape, texture, color, and so on, since relatively identical objects in general are not always indiscernible in these respects. I thus conclude that Deutsch's principle is too coarse-grained for the purposes at hand, in that it fails to yield a satisfying explanation for the striking similarities that are conferred upon objects by the various identity-like relations collected under the heading "relative identity."

III. CONCLUSION

This paper examined a variety of contexts in metaphysics which employ a strategy I consider to be suspect. In each of these contexts, the Suspect Strategy (TSS) aims at excluding a series of troublesome contexts from a general principle whose truth the philosopher in question wishes to preserve. Our main representatives of (TSS), as applied to Leibniz's Law, (LL), or restricted versions thereof of the form, (RI), were Gibbard's defense of contingent identity, Parsons's defense of indeterminate identity, as well as Deutsch's recent development of the relative identity view. On the basis of these examples, we discerned four different ways in which (TSS) can be implemented: (i) the most widespread Purely Stipulative Strategy; (ii) Gibbard's Appeal to Failures of Substitutivity; (iii) Parsons's Appeal to the Paradoxes of Naive Set Theory; and (iv) Deutsch's Expansion Principle. I discussed in detail why I believe that (TSS) remains suspect in all four of the approaches considered above.

While we of course cannot conclude from our exposure to extant versions of (TSS) that *no* exclusion procedure could ever overcome the troubling features we encountered, my remarks here should, I think, at least give us reasons to be skeptical that any strategy which proceeds by means of *purely formal* (for example, syntactic) individuation criteria could achieve its intended purpose; for we have seen that such strategies in general tend to be *too coarse-grained* to individuate contexts correctly into those that should and those that should not be excluded from the reaches of the general principle under discussion. I suspect, moreover, though I did not argue for this stronger claim, that any strategy which does *not* proceed by means of purely formal criteria would in some way succumb to the charge of *circularity*.

Supposing that no nonsuspect strategy can be found by which to exclude the troublesome contexts from the reaches of the general principle, where does this leave us? As I see it, we have basically two options: (i) we can either accept that the general principle in question is true, that the relevant contexts denote properties and that these properties fall under the scope of the general principle; or (ii) we can deny the truth of the general principle in question. The second option, I take it, is not one that many philosophers would take seriously in the context of (LL) or certain instances of (RI), but it may be one that is attractive in other cases (for example, that of Parsons's existence principle, (EP)). If the truth of the general principle is nonnegotiable, then option (i), in the absence of further independently motivated considerations, naturally leads to a universe populated with a surprising multitude of numerically distinct yet almost indiscernible objects, such as statues and the lumps of clay that constitute them. For, assuming the preceeding remarks are correct, (TSS) now can no longer be invoked in order to bracket those contexts, such as '____ is essentially a statue', by means of which these objects are apparently discernible; and objects which are merely *almost* indiscernible, by Leibniz's Law, are numerically distinct.

KATHRIN KOSLICKI

Tufts University

HOW TO READ "EPISTEMOLOGY NATURALIZED"*

> W.V. Quine is well known for urging the abandonment of epistemology, as traditionally pursued, in favor of the scientific project he calls "naturalized epistemology."[1]

> Epistemology, on Quine's view, is a branch of natural science;...[it] should not concern itself with how we should form beliefs but with how we do in fact form them.[2]

> [Quine] is generally credited with having started the move toward naturalism in epistemology.... Quine's view appears to leave no room for normative evaluations in epistemology.[3]

No one, to my knowledge, would dispute these characterizations of the burden of Quine's "Epistemology Naturalized"[4] and its role in "the move toward naturalism in epistemology." Yet every student of Quine knows that as early as "Two Dogmas of Empiricism"[5] and as late as *Pursuit of Truth*[6] he pursued normative epistemological investigations that cannot sensibly be seen as falling within the purview of natural science or, more particularly, psychology. Hilary Putnam, the only philosopher I know to have considered how this glaring conflict might be resolved, was forced to throw up his hands; although Quine repeatedly assured him that he did not mean to "rule out the normative," Putnam could find no way of reading "Epistemology Naturalized" except as a proposal to do exactly that.[7] Eventually,

---

* I dedicate this essay to the memory of Willard van Orman Quine.

I am indebted to Roger Wertheimer, as always, for much discussion, and for giving me the benefit of his unerring critical eye; I thank Paul Saka, as well, for perceptive comments on two late drafts. My work was supported by a Faculty Development Leave from the University of Houston's College of Liberal Arts and Social Sciences.

[1] Ernest Sosa and Jaegwon Kim, eds., *Epistemology: An Anthology* (Malden, MA: Blackwell, 2000), p. 289.

[2] Sven Bernecker and Fred Dretske, eds., *Knowledge: Readings in Contemporary Epistemology* (New York: Oxford, 2000), p. 233.

[3] John L. Pollock and Joseph Cruz, eds., *Contemporary Theories of Knowledge* (Lanham, MD: Rowman and Littlefield, 1999, 2d ed.), p. 165.

[4] In Quine's *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 69–90; hereafter EN.

[5] In Quine's *From a Logical Point of View* (Cambridge: Harvard, 1961, rev. ed.), pp. 20–46.

[6] Cambridge: Harvard, 1992, rev. ed.

[7] Putnam, "Why Reason Can't Be Naturalized," Howison Lecture, University of California, 1981; published in *Realism and Reason*, volume 3 of his *Philosophical Papers* (New York: Cambridge, 1983), pp. 229–47.

0022-362X/05/0202/78–93

Quine gave the world at large the same assurance, and even sketched his views of the fundamental norms of epistemology and how they are rightly arrived at,[8] but no one has yet done better than Putnam in seeing how the essay can be squared with that assurance or with Quine's undiminished interest in normative questions.

I take it as a given that the universal reading of "Epistemology Naturalized" as a proposal to "surrender the epistemological burden to psychology" is *wholly* mistaken, since the alternative—that Quine, in mid-career and without so much as a "by your leave," delivered himself of a proposal whose implementation would have required him to abandon some of his deepest lifelong concerns, and that he failed for decades to detect the conflict—is utterly absurd. Here I shall offer a reading of the essay on which its central proposal is consistent with his fundamental views, and briefly discuss a few related matters. On my reading, not only does Quine not urge the abandonment of epistemology as traditionally pursued, but, whatever inspirational role "Epistemology Naturalized" may have played in the move toward "naturalism in epistemology," its central proposal is actually antithetical to a good deal of what is nowadays done under that heading.

The universal misreading of Quine's essay is unquestionably a bizarre phenomenon, and so, fittingly, is its explanation: the fault lies not in the universal incompetence of the essay's readers, but rather in a disastrous failure of its author's outsized gifts as an expositor of his own views. No one familiar with Quine's work could arrive lightly at such an assessment of any part of it, nor have I done so in this case; in trying to understand how things went wrong I have, however, been buoyed by one truth I take to be beyond peradventure: better expository failure than perdurable blindness to blatant conflict between principle and practice.

We begin at ground zero. After noting some milestones in the history of attempts to "account for the external world as a logical construct of sense data," including Rudolf Carnap's efforts at rational reconstruction, which, if successful, "would have enabled us to translate all sentences about the world into terms of sense data, or observation, plus logic and set theory," Quine asked his now famous questions:

> [W]hy all this creative reconstruction, all this make-believe? The stimulation of his sensory receptors is all the evidence anybody has had to go

---

[8] See Quine, "Reply to Morton White," in Lewis Edwin Hahn and Paul Arthur Schilpp. eds., *The Philosophy of W.V. Quine* (LaSalle, IL: Open Court, 1986), pp. 664–65, and *Pursuit of Truth*, p. 19. I will return to this point later.

> on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology (EN 75)?

Here Quine muddies the water at the very outset. The reader could easily take him to be contrasting two projects with identical starting points—a subject's experiences—and equivalent endpoints: two versions of a single picture of the world. That is, he might seem to be suggesting that there is a way, not necessarily "rational," in which subjects *actually* construct their pictures of the world from their experiences, and a way, sketched by Carnap, in which those pictures *could be rationally (re)constructed* from those same experiences, and proposing that we settle for the former. But Carnap's project was *not* to construct rational routes from a subject's data to his picture of the world; rather, it was to devise (re)constructions of our ordinary physicalistic *concepts* that would allow us to "translate all sentences about the world into terms of sense data, or observation, plus logic and set theory." Even if his project had succeeded, no subject would have been in a position to infer the Carnapian translation of a typical physicalistic sentence from his experiences by tracing some Carnapian path from the one to the other. One reason is elementary: the translation of a typical sentence about the world would have referred to vastly more observations than any subject could possibly make, a point Quine illustrated by noting that "The most modest of generalizations about observable traits will cover more cases than its utterer can have had occasion actually to observe" (EN 74). A deeper and more theoretically illuminating reason is that *there could not in principle be* any such translations of most sentences, since "the typical statement about bodies has no fund of experiential implications it can call its own" (EN 79).

In the terms of Quine's broad framework in "Epistemology Naturalized," subjects' constructions of their world-pictures and Carnap's efforts at rational reconstruction fall on opposite sides of his division of epistemology into doctrinal and conceptual studies. Subjects engaged in constructing pictures of the world are developing "doctrines" that are the objects of *doctrinal* studies, which concern, *inter alia*, the justifiability of those doctrines; Carnap's efforts at reconstruction are *conceptual* studies aimed at showing that the everyday concepts employed in *formulating* those doctrines are theoretically dispensable in favor of sense data, or observations, plus logic and set theory.

But if Quine is not contrasting the individual subject's actual construction of his picture with a hypothetical Carnapian parallel, then what *is* he contrasting it with? What, exactly, is the study he proposes to abandon in favor of the study of psychology? The answer to these questions should have been unmistakable, though everyone seems in fact to have mistaken it: "all this creative reconstruction, all this make-believe," that is, the conceptual studies pursued by Carnap and others. This means that the question—"Why not settle for psychology [rather than pursue the project of creative reconstruction]?"—is quite peculiar; it would make clear rhetorical sense only if the two projects were at least roughly parallel, or comparable. That Quine should have posed this odd question is itself mildly unfortunate, but in doing so he also set the stage for something much worse: the principal expository blunder responsible for the misreading of his essay. Having proposed to settle for psychology in place of "creative (rational) reconstruction," he went on to *describe* such a settling as "a surrender of *the epistemological burden* to psychology" (EN 75; my emphasis); given that description, his readers can hardly be faulted for having read him as proposing to abandon *epistemology itself* in favor of psychology.

Nevertheless, that reading is as mistaken as the description is misleading. Note first that the description, as it would be understood by any well-informed reader, is absurd; while epistemology *includes* efforts at rational reconstruction, it certainly does not *consist of* them, so that abandoning them would hardly amount to abandoning the entire discipline. The extraordinary description is explained by an equally extraordinary fact: as Quine uses the term 'epistemology' in "Epistemology Naturalized," that discipline *does* consist of efforts at rational reconstruction—or, more broadly, attempts to discover or establish logical, translational or rational-reconstructive links between observation and theory. This is readily confirmed simply by rereading the essay; I shall cite only two passages in whose conjunction it is starkly evident:

> The old epistemology aspired to...*construct* [natural science] somehow from sense data (EN 83; my emphasis).

> Philosophers have rightly despaired of *translating* everything into observational and logico-mathematical terms.... And some philosophers have seen in this *irreducibility* the bankruptcy of *epistemology*....
>
> But I think that at this point it may be more useful to say rather that epistemology still goes on.... Epistemology, or something like it, simply falls into place as *a chapter of psychology*... (EN 82; my emphases).

In short, discovering the impossibility of reduction, or translational reconstruction, *would have* meant the bankruptcy of epistemology were it not that it could be *reconceived* as a chapter of psychology.

Quine's idiosyncratic use of "epistemology" explains another oddity as well. In briefly sketching a few salient points in the history of modern epistemology, he restricts his attention entirely to efforts of a logical or quasi-logical sort by philosophers from Hume to Carnap; one will search in vain, in "Epistemology Naturalized" or elsewhere in his writings, for any substantial discussion of what might be called "mainstream" efforts by numerous epistemologists to show that belief in the external world is, or can be, justified in the *absence* of such logical or quasi-logical links. Indeed, he appears at times to be entirely oblivious of those efforts:

> In *the old epistemological context* the conscious form [of observation, namely, conscious three-dimensional apprehension] had priority, for we were out to *justify our knowledge of the external world by rational reconstruction*, and that demands awareness. Awareness *ceased* to be demanded when we *gave up* trying to justify our knowledge of the external world by rational reconstruction (EN 84; my emphases).

The suggestion that awareness is not demanded by any conceivable effort *other* than rational reconstruction to justify our belief in the external world would have come as news to those many epistemologists who thought (as many still do) both that some role for awareness is demanded for epistemic justification and that such justification is possible in the *absence* of rational-reconstructive links between evidence and theory.[9]

Given, then, that what Quine proposes to abandon is only the project of rational reconstruction, we can reformulate his stage-setting questions as follows: "Why all this searching for logical or quasi-logical connections between evidence and theory? Why not just see how we construct our theories from our evidence? Why not settle for psychology?" An apparent lacuna in "Epistemology Naturalized," one that prompted a second question from Putnam, is made even more glaring by this reformulation: Why did Quine not consider the possibility of finding an *inductive* evidence/theory connection, once he became convinced that no deductive, or quasi-deductive, one was to be found? But philosophers puzzled by this (as I certainly was) have presumably failed to register the fact that, as early as "Two Dogmas of Empiricism," he had rejected the idea that any such connection existed:

> [Physical objects are *myths* on the same footing with abstract entities]

---

[9] This is the more remarkable in light of the fact that Quine counted C.I. Lewis and Roderick Firth, two of the century's staunchest defenders of classical foundationalism, among his Harvard colleagues for many years.

> and gods, *neither better nor worse except for* differences in the degree to which they expedite our dealings with sense experiences.[10]

It follows from this—and I take it that only Quine's distaste for redundancy kept him from making the point explicitly—that the myth of physical objects is no more *probable* relative to our evidence (whether experiences, sense data, observations, or stimulations), is no more *justified by* that evidence, than is the myth of the Homeric gods; the *only* relevant difference between the two is that the former better expedites our dealings with sense experiences.

In *Word and Object* he made the point explicitly:

> [T]he statements ['Brutus killed Caesar' and 'The atomic weight of sodium is 23']...are justifiable *only* by *supplementing* observation with scientific method....[11]

To indulge again in redundancy, observations alone cannot justify theories.

His clearest and most emphatic statement of the point came in *Pursuit of Truth,* in one of the few places where he can sensibly be read as at least alluding to foundationalists' contentions that our experience provides support for our belief in an external world:

> Traditional epistemology sought grounds in sensory experience capable of implying our theories about the world, or at least of endowing those theories with some increment of probability. Sir Karl Popper has long stressed, to the contrary, that observation serves only to refute theory and not to support it. We have...been seeing in a schematic way why this is so.
>
> ...It is clearly true...that one continually reasons not only in refutation of hypotheses but in support of them. This, however, is a matter of arguing logically or probabilistically from other beliefs already held.... Some of those supporting beliefs may be observational, but they contribute support *only in company with* others that are theoretical. *Pure observation lends only negative* evidence....[12]

---

[10] "Two Dogmas of Empiricism," p. 45; my emphases. Perhaps it is worth noting that "expediting our dealings with sense experiences" is not solely a matter of enabling us to get the predictions right; considerations of simplicity, economy, and so forth also enter in.

[11] *Word and Object* (Cambridge: MIT, 1960), p. 24; my emphases. For a defense of the idea that this was the heart of Hume's skeptical argument concerning inductive inference, see Colin Howson, *Hume's Problem* (New York: Oxford, 2000), pp. 10–21.

[12] *Pursuit of Truth,* pp. 12–13; my emphases.

How is this position to be squared with Quine's numerous references in "Epistemology Naturalized" to sensory or observational evidence, including his remark that "[W]hatever evidence there is for science is sensory evidence" (EN 75)? Given that supporting evidence, in general, enhances a theory's probability, such remarks seem clearly to imply that relevant observations, in particular, do so.

Calcutta University ALIPORE CAMPUS

Quine's position, then, is that there are neither deductive, nor rational-reconstructive *nor inductive* links to be found between evidence and theory, and that if we want to understand how the two are related we should settle for psychology—we should study how *we relate* the two, that is, how we actually construct our theories from our evidence.[13]

Before considering where he thinks such an investigation would lead, we need to address an obvious preliminary question about its viability. Quine is clearly proposing that we study *the* way subjects construct their pictures of the world, not countless idiosyncratic ways in which individual subjects may do so. But how could he sensibly suppose that there is a single such way, given that subjects with comparable stimulation histories construct widely varying, even conflicting, world pictures?

Quine could do so because he thought he had long since identified that way. In his view, our constructions have two stages, each characterized by a uniform method: in the first stage we construct our proprietary versions of "the immemorial doctrine of ordinary enduring middle-sized physical objects"[14] (our "doctrines of an external world"); in

---

First, and most importantly, all of the remarks in question are perfectly compatible with his contention that sensory evidence enhances the probability of theory *only in conjunction with* (other) theory. Second, the fact that the issue is at best tangential to the central thrust of "Epistemology Naturalized" may well have been reason enough for Quine, that paragon of expository economy and elegance, to allow it to remain unstated. Indeed, it is presumably such considerations that explain how even here (*Pursuit of Truth*), where the Popperian point is at center stage, he could write the following:

> We were undertaking to examine the evidential support of science. That support, by whatever name, comes now to be seen as a relation of stimulation to scientific theory....
>
> Observation sentences are thus the vehicle of scientific evidence, we might say.... Insofar as observation sentences bear on science at all, affording evidence...(pp. 2–7).

Finally, there is of course his never-abandoned view that a theory positing physical objects is epistemologically superior to one positing gods *only* in contributing more effectively to the prediction of experience.

[13] We should perhaps note Quine's suggestion that "something like...rational reconstruction" might still be worth pursuing for heuristic reasons—"imaginative constructions can afford hints of actual psychological processes" (EN 83)—and the possibility that he continued to think, with Nelson Goodman and others, that efforts at partial reconstruction can be of intrinsic interest. See Quine's "On What There Is," *Review of Metaphysics*, II (1948): 21–38, reprinted in *From a Logical Point of View* (New York: Harper and Row, 1961), pp. 1–19, and Goodman's *The Structure of Appearance* (Cambridge: Harvard, 1951), along with sections 1–3 of his "The Revision of Philosophy," in Sidney Hook, ed., *American Philosophers at Work* (New York: Criterion Books, 1956), reprinted in his *Problems and Projects* (Indianapolis: Bobbs-Merrill, 1972), pp. 5–23.

[14] *Word and Object*, p. 11.

the second we construct our (more or less) comprehensive theories about that world.

Relative to our interests, the most significant difference between the two stages is that, unlike our comprehensive theories, our doctrines of an external world are essentially uniform, so that there is no obvious difficulty with the idea that we have a common way of constructing them. Not only did Quine suppose that there was such a way, but from *Word and Object* to *Pursuit of Truth* he speculated brilliantly and in some detail about a number of its stages.

Our principal interest lies in understanding his view of our second-stage constructions, which turns out to be the key to understanding why his proposal to naturalize epistemology does not involve abandoning the normative. It is clearly these constructions that he has in mind when he says that his aim is "to understand the link between observation and *science*" (EN 76, my emphasis), or that "We are studying how the human subject of our study posits bodies and projects his *physics* from his data" (EN 83, my emphasis), or that "There is thus reciprocal containment...: epistemology in science and *natural science* in epistemology" (EN 83, my emphasis); and it is here that we reencounter our original problem: How can Quine suppose that we have a uniform way of constructing these theories, given that conflict among them is rife?

There is a second problem here as well. Quine is quite naturally read as suggesting that, for purposes of studying the relation between observation and theory, subjects' relevant output consists of their first-order scientific theories. But in many cases that output also includes second-order theories concerning the assessment of first-order ones, and three things are worth noting about these metatheories: they embody their subjects' own views about the very relation between evidence and theory that we are trying to understand; they play a role—along with observation—in determining which first-order theories those subjects hold; and they, too, will vary widely, even among subjects with comparable stimulation histories. Though taking into account this additional output might not render the proposed project incoherent, it would certainly make it considerably more challenging, even without considering the specter of yet higher levels of theory. (Imagine that our subjects included theorists from Bacon, Bayes, and Mill to Karl Popper, Bruno deFinetti, and Nelson Goodman.)

Given these difficulties with the idea that we share a common method of second-stage construction, we might ask how Quine himself attempted to identify it. Did he, in accordance with his own apparent proposal in "Epistemology Naturalized," implement any new psychological investigation? Never; neither there nor in any of his subsequent writings did he initiate any new line of inquiry, psychological or otherwise, that could conceivably be seen as an attempt to identify that method. What are we to make of this? How are we to understand his apparent failure to take up his own proposal? These questions are wrong headed, and simply evaporate once we realize that "Epistemology Naturalized" does not contain any such proposal. Quine's aim in that essay was not to launch a new line of investigation, but rather to persuade his readers to join in an *old* one in which he had long been engaged. If we want to locate the *origins* of his efforts to naturalize epistemology, we have to look *backward* from "Epistemology Naturalized."

At least as early as *Word and Object*, Quine had carefully considered the relation between subjects' inputs and outputs, and discovered a good deal about how those subjects relate evidence to theory:

> [S]implicity considerations in some sense may be said to determine even the least inquisitive observer's most casual acts of individual recognition. For he is continually having to decide, if only implicitly, whether to construe two particular encounters as repeated encounters with an identical physical object or as encounters with two distinct physical objects. And he decides in such a way as to minimize, to the best of his unconscious ability, such factors as multiplicity of objects, swiftness of interim change of quality and position, and, in general, irregularity of natural law.
>
> ...The deliberate scientist goes on in essentially the same way...and a law of least action remains prominent among his guiding principles. Working standards of simplicity, however difficult still of formulation, figure ever more explicitly.[15]

Reflection on the most elementary facts about the nature of subjects' inputs and outputs revealed that even the least inquisitive among us conform in our theorizing to some version, or proto-version, of scientific method; it revealed that our second-stage method is that of constructing theory from evidence in conformity with a norm of maximizing simplicity.

Is this, then, at least a broad-brush, preliminary version of Quine's proposed project and the knowledge to be acquired by pursuing it? On the universal reading of "Epistemology Naturalized," it is; in the language of the quotations with which this paper began, the project described is certainly a "scientific" one, it does "not concern itself with how we should form beliefs, but with how we do in fact form them," and it seems to have no place for "normative evaluations"—it issues simply in a description of how we do in fact construct our theories from our evidence. But if Quine's idea were simply that we should abandon normative questions in favor of scientific ones, then why psychology in particular? Why not, say, geology? There must be *some* significant relation between rational reconstruction and psychology if the suggestion that we "settle for" the latter in place of the former is to be at all sensible, and indeed there is: both promise to illuminate the relation between evidence and theory. It is Quine's revolutionary conviction that by learning the psychological truth that *we relate* evidence to theory in conformity with scientific method, we learn the philosophical truth that evidence *is related* to theory by scientific method. This is a point he had long since advertised:

---

[15] *Word and Object*, p. 19. Here Quine mentions only simplicity, or "least action," considerations as entering into theory choice, but elsewhere he notes other "systematic" considerations—notably conservatism (conserving as much as possible of one's existing body of beliefs in revising it to accommodate recalcitrant data), but also generality, refutability, and modesty (roughly, limiting one's ontological commitments to the minimum needed to account for one's data). See "Hypothesis," chapter 5 of Quine and J.S. Ullian, *The Web of Belief* (New York: Random House, 1970),

> [S]cientific method, whatever its details, produces theory whose connection with all possible surface irritation [evidence] consists solely in scientific method itself, unsupported by ulterior controls.[16]

To put his point even more starkly, *the only connections there are* between our evidence and our theories are *those we make* in accordance with scientific method. (To avoid any suggestion that this method is independent of us—something we have discovered, rather than devised—it might better have been called "our" scientific method, but for the latter's suggestion that it is one among others.)

But how, then, can Quine claim that by learning how *we relate* evidence to theory we learn how evidence *is related* to theory? He can do so because in his view the former determines the latter, which is a corollary of his more general view that there is no exile from theory. By the latter he means, *inter alia,* that there is no vantage point from which our comprehensive method of devising theories to explain past observations and predict future ones can be evaluated in the relevant respect: we have no independent way of identifying true theories that would allow us to consider whether our method of theorizing tends, or tends better than would alternative methods, to yield true theories: "[W]e can never do better than occupy the standpoint of some theory or other, the best we can muster at the time."[17] In terms of Quine's favorite simile, borrowed from Otto Neurath, our comprehensive theory is like a boat at sea, and we have no alternative to staying on board if we want to stay afloat, that is, to think rationally at all; if we want to find out how to arrive at the truth on the basis of our evidence—if, that is, we want to understand how evidence relates to theory—we have no alternative but to take (our best understanding of) the way we do arrive at theory as the way we are to arrive at theory. It is, of course, an all-important fact that any part of our comprehensive theory—including any statement about how evidence relates to theory—can be modified or repaired, that "no statement [belief] is immune to revision":[18]

---

[16] *Word and Object,* p. 23.

[17] *Word and Object,* p. 22.

> [W]e own and use our beliefs of the moment, even in the midst of philosophizing, until by what is vaguely called scientific method we change them here and there for the better.[19]

We have now arrived at an entirely new picture of Quine's principal efforts in "Epistemology Naturalized"; so far is he from proposing to abandon the normative that he is proposing instead to *discover* the norms that govern theorizing by discovering the norms that we conform to in our theorizing. This means that his talk of surrendering the burden of epistemology to psychology is unfortunate in a second respect as well. What he is actually proposing is to *enlist the aid* of psychology in addressing the burden of epistemology: psychology will identify the norms we adhere to, and philosophy will tell us that, *by virtue of* their being the ones we adhere to, they are the ones we *are to* adhere to. He himself later put the point in much the same way:

> The channels by which, having learned observation statements, we acquire theoretical language, are the very channels by which observation lends evidence to scientific theory....
>
> We see, then, a strategy for investigating the relation of evidential support between observation and scientific theory. We can adopt a genetic approach, studying how theoretical language is learned. For *the evidential relation is virtually enacted*, it would seem, in the *learning*.[20]

Several loose ends need tidying up. First, the question, "Why not settle for psychology?" is at best a bit peculiar if we have no alternative—if, that is, there are no connections between observation and theory other than those we make, hence none for us to discover—and I can see no way of reading it so as to obviate this peculiarity altogether.

---

[18] "Two Dogmas of Empiricism," p. 43; my emphasis.

[19] *Word and Object*, pp. 24–25.

[20] Quine, "The Nature of Natural Knowledge," in Samuel Guttenplan, ed., *Mind and Language* (New York: Oxford, 1975), pp. 67–81; quote from pp. 74–75, my emphases.

Yet there is an important point to be made about it. We have seen that for Quine settling for psychology does not mean settling for mere description; getting the relevant psychological descriptions right is only the first step in arriving at an understanding of the norms of epistemology. Thus I suggest that Quine may have had in mind something like the following question: "Why not settle for the justification that we can see we already enjoy, rather than search for a further justification via rational reconstruction (especially since none is to be had)?" This question captures the central thrust of his essay, to wit, that we should stop beating a dead horse—stop trying to devise some ingenious story about how our beliefs are justified by virtue of some intrinsic relation in which they stand to our evidence—and remind ourselves that the only possible account of epistemic justification, hence the one for which we should "settle," is one framed in terms of the scientific method to which psychology tells us we conform, which simply *is* "the way to truth."[21]

A second loose end concerns the original apparent obstacle to the idea that there is a single way in which we construct our comprehensive theories: If we all construct them in accordance with scientific method, then what explains the fact that subjects with comparable stimulation histories arrive at conflicting theories? The explanation is that, although we all rely on the scientific method, we arrive at different judgments about *which theories are sanctioned* by that method. The relevant principle of scientific method runs something like this: *ceteris paribus,* Theory T is superior to Theory T' if it is simpler, or more general, or stronger than T' relative to the combination, or some weighted combination, of the various "systematic virtues," *or* adopting T would require abandoning, or modifying, fewer, or less fundamental, previously held beliefs than would adopting T', *or....* Not only is there an enormous amount of "slack" in this "principle," but we differ in our weightings of the various systematic virtues, and these considerations might well suffice to account even for ideally rational agents' arriving at conflicting theories on the basis of comparable bodies of evidence.[22]

Finally, I want to consider briefly the passages mentioned earlier in which Quine offered his assurances that he was not proposing to abandon the normative, since they contain what might seem to be yet another puzzle about his conception of naturalized epistemology:

---

[21] *Word and Object,* p. 23.

[22] For a striking illustration of individual differences in weighting, see Freeman Dyson's sketch of Einstein's and Poincaré's differing degrees of receptivity to radical theory change in his "Clockwork Science," *New York Review of Books* (November 6, 2003): 42–44.

> Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in a more cautiously epistemological term, prediction. Like any technology, it makes free use of whatever scientific findings may suit its purpose.[23]

> [N]aturalized epistemology on its normative side is occupied with heuristics generally—with the whole strategy of rational conjecture in the framing of scientific hypotheses.[24]

What is new and puzzling here is Quine's use of the term 'engineering', and his apparent suggestion that while engineering is normative (in that it attempts to construct the best device for a given purpose), science is descriptive (and provides engineers with its "findings"). But compare this with two of many similar antecedents in his writings:

> As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience.[25]

> A sentence's claim to scientific status rests on what it contributes to a theory whose checkpoints are in prediction.[26]

Here *scientific theories* yield predictions. Of course, in *deriving* those predictions we are constrained by logical norms, but this is obviously not what Quine has in mind by the "technology" of prediction. What has happened, I suggest, is simply that he has adopted new terminology for making familiar points. What he here calls engineering is simply theorizing—thought of now as devising/constructing scientific theories designed to maximize predictive efficacy. As for the scientific "findings" made use of in such projects, they are simply whatever theories we take for the nonce to have been established—in accordance with our normative scientific method—along with our observations, especially those that are currently inexplicable or have falsified earlier theories. On this reading, normative epistemology does indeed fall into place as a part of science; its distinctive tasks are to identify, formulate, evaluate, and perhaps modify the most general norms that guide our scientific theorizing, and to evaluate our theories in light of those norms; even in its new clothing and setting, this is unmistakably epistemology as traditionally pursued.

---

[23] "Reply to Morton White;" pp. 664–65.

[24] *Pursuit of Truth*, p. 19.

[25] "Two Dogmas of Empiricism," p. 44; my emphasis.

[26] *Pursuit of Truth*, p. 20.

Given my proposed reading of "Epistemology Naturalized," several points about the relationship between epistemology and psychology, and Quine's view of it, deserve brief consideration.

Despite its reputation, the essay contains little that is new; it is largely a sustained argument for the idea that much if not all of what goes on under the name of epistemology can be seen to be part of the scientific enterprise broadly—and rightly—conceived. We have seen, however, that, even so conceived, Quine's proposal about how the nature of the evidence/theory relation should be determined is best described as a matter of enlisting the aid of psychology in that investigation. This is certainly a striking example of the occasional direct relevance of psychology to philosophy, and in this it resembles John Rawls's reliance, in developing his theory of justice, on psychological hypotheses about what the agents in his "original position" would do; but neither Quine nor Rawls is aptly described as having surrendered a philosophical burden to psychology.[27]

At the conclusion of his essay, Quine mentions several theoretical endeavors that, he suggests, might be furthered by "rubbing out the boundaries" between epistemology and science (psychology and linguistics, in particular), but his suggestions are quite puzzling. He mentions the possibilities of (i) determining whether there is an "alphabet" of perceptual norms toward which we tend unconsciously to rectify all perceptions, which "could be taken as epistemological building blocks, the working elements of experience," (ii) determining whether various features of our cognitive processes can be explained on the basis of their survival value, and (iii) "clarifying" induction by appeal to evolution. I simply do not understand what he can have in mind in suggesting that (i) or (ii), as interesting as they are, might be furthered by rubbing out disciplinary boundaries, though (i) might be another case of psychology's being an important aid to an epistemological investigation. As for (iii), Quine himself once attempted a justification of induction by appeal to evolutionary theory, but abandoned it (whether rightly or wrongly does not matter here) in the face of Goodman's new riddle of induction, and later described a different role for that theory:

> [Appealing to Darwinian biology to justify induction] would be circular,

---

[27] See Rawls, *A Theory of Justice* (Cambridge: Harvard, 1974), especially chapter III, sections 24–26.

> since biological knowledge depends on induction.... I am granting the efficacy of induction, and then observing that Darwinian biology, if true, helps explain why induction is as efficacious as it is.[28]

There is nothing to dispute here, but neither is there anything to suggest that rubbing out boundaries promises theoretical progress. Of course Quine may have had in mind something else altogether, but I am at a loss to imagine what that might be. In sum, it is hard to see in his brief remarks any promise of theoretical progress to be made as a result of eliminating this boundary.[29]

Finally, while many, perhaps most, philosophers have seen in "Epistemology Naturalized" the original impetus toward various views now often grouped under the title "naturalized epistemology," and while various such views were doubtless inspired by that essay as it has hitherto been read, not only do some of them have no discernible relation to the proposals actually contained in it, but Quine would have objected vigorously to them.

Given his enormous influence on epistemology, it can come as something of a surprise to realize that he had virtually no interest in much of the work that held center stage in that discipline during his lifetime. We have already noted his lack of interest in mainstream traditional theories, and two related phenomena deserve highlighting: when he suggested ways in which progress might result from rubbing out the boundary between epistemology and psychology, he did not so much as hint at the possible development of "naturalistic" or "psychologistic" accounts of either knowledge or justification, and in later years he showed no interest at all in the proliferating "externalist" accounts of those concepts.

Concerning knowledge, or what he once referred to as "our so-called knowledge,"[30] he had next to nothing to say; such little thought as he gave it may well have been captured in his response to the question, "How can we know that [one complete physical theory] is true and [another] false?":

> There is an obstacle in the verb "know." Must it imply certainty, infallibility? Then the answer is that we cannot. But if we ask rather how we are

---

[28] "The Nature of Natural Knowledge," p. 70.

[29] Earlier in the essay, Quine suggests that one result of "seeing epistemology in a psychological setting" (which may or may not amount to rubbing out the boundary between the two) is that "What to count as observation now can be settled in terms of the stimulation of sensory receptors, let consciousness fall where it may" (EN 84). Certainly it *could* be, but it is by no means obvious that it *must* be; this matter, too, seems to be independent of the issue of disciplinary boundaries.

[30] "Two Dogmas of Empiricism," p. 42.

> better warranted in believing one theory than another, our question is a substantial one. A full answer would be a full theory of observational evidence and scientific method.[31]

There is no reason to suppose that Quine had any interest whatever in the concept of knowledge, much less that he saw his proposals in "Epistemology Naturalized" as bearing on its proper explication.

As for justification, we have already seen one perfectly adequate explanation for his lack of interest in contemporaneous accounts of it: he thought he had already specified the right account in terms of scientific method. However, there is an equally important, and deeply ironic, point to be made about how he would have viewed externalist accounts of it framed in terms of such things as reliable processes and causal relations: *he would have faulted them for abandoning the normative in favor of the descriptive.*

BREDO C. JOHNSEN

University of Houston

---

[31] Quine, *Theories and Things* (Cambridge: Harvard, 1981), p. 180.

COMMENTS AND CRITICISM

THE MEANING OF 'RACE': FOLK CONCEPTIONS AND THE NEW BIOLOGY OF RACE*

The idea that race is biologically real has played a central role in scientific discussions of race, and in *common sense* (CS) as well, for some time. Indeed, some historians argue that scientific and folk conceptions of race developed hand in hand around the late eighteenth century.[1] While a number of biological definitions have been proposed, many share the assumption that it is possible to divide humans into races such that the members of each share certain heritable characteristics, such as overt physical traits and some psychological or behavioral traits as well, that they do not share with members of other races. Though this assumption went largely unquestioned for some time, today the problems with it are fairly well known. Many of the traits typically used to individuate races vary independently.[2] Moreover, research in human genetics shows that there is more genetic variation within than among the three major races.[3] Most academics have taken these problems to show that race is biologically unreal, but Philip Kitcher and I take a different view.[4] We have independently argued that while the above assumption is problematic, most problems faced by earlier biological conceptions can be avoided by defining race genealogically. Kitcher and I are not alone in defining race this way; there is a growing number of biologists who endorse similar definitions.[5]

---

* I thank Fred Adams, Kai Draper, Richard Hanley, and Doug Stalker for their helpful comments.

[1] Michael Banton and Jonathan Harwood, *The Race Concept* (London: David & Charles, 1975); Audrey Smedley, *Race in North America* (Boulder: Westview, 1993); David Goldberg, *Racist Culture* (Cambridge: Blackwell, 1993); Anthony Appiah, "Race, Culture, Identity," in *Color Conscious* (Princeton: University Press, 1996), pp. 30–105.

[2] Frank Livingstone, "On the Nonexistence of Human Races," *Current Anthropology*, III, 3 (June 1962): 279–81.

[3] Richard Lewontin, "The Apportionment of Human Diversity," in Theodosius Dobzhansky et al., eds, *Evolutionary Biology 6* (New York: Appleton-Century-Crofts, 1972), pp. 381–98.

[4] Andreasen, "A New Perspective on the Race Debate," *British Journal for the Philosophy of Science*, XLIX, 2 (June 1998): 199–225; Andreasen, "Race: Biological Reality or Social Construct?" *Philosophy of Science*, LXVII, 3S (September 2000): S653–66; Andreasen, "The Cladistic Race Concept: A Defense," *Biology and Philosophy*, XIX (June 2004): 425–42; Kitcher, "Race, Ethnicity, Biology, Culture," in Leonard Harris, ed., *Racism* (Amherst, NY: Humanity, 1999), pp. 87–120.

[5] Masatoshi Nei and Arun Roychoudhury, "Evolutionary Relationships of Human

0022-362X/05/0202/94–106

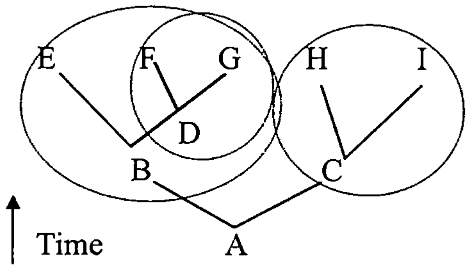


Figure 1: In this figure, the letters represent well-defined species, the branches represent speciation events, and the circles identify monophyletic groups.

I have developed what I call *the cladistic race concept* (CRC). Cladistics is a branch of systematic biology that defines taxa solely in terms of common ancestry. Cladistic classification begins with a well-confirmed evolutionary branching structure (see figure 1) and defines taxa as monophyletic groups (groups that are composed of an ancestor and all of their descendents).[6] Typically, cladistic classification is used for defining higher taxa. With reference to a diagram depicting evolutionary relations among species, higher taxa are defined monophyletically. Nonetheless it is possible to apply these ideas to race. A number of research groups have proposed branching diagrams (see for example, figure 2) that aim to represent evolutionary relations among reasonably reproductively isolated human breeding populations.[7] I maintain that races can be defined as monophyletic groups of such populations.

As already mentioned, Kitcher also argues that races can be defined,

---

Populations on a Global Scale," *Molecular Biology and Evolution*, x, 5 (1993): 927–43; H. B. Shaffer and Mark McKnight, "Polytypic Species Revisited," *Evolution*, L, 1 (1996): 417–33; John Legge et al., "Genetic Criteria for Establishing Evolutionarily Significant Units in Cryans Buckmoth," *Conservation Biology*, x, 1 (1996): 85–98; Neil Risch et al., "Categorization of Humans in Biomedical Research: Genes, Race and Disease," *Genome Biology*, III, 7 (September 2002): 1–12; Noah Rosenberg et al., "Genetic Structure of Human Populations," *Science*, CCXCVIII (December 2002): 2381–85; Michael Bamshad and Steve Olson, "Does Race Exist," *Scientific American*, CCLXXXIX, 6 (December 2003): 78–85.

[6] It is important not to confuse cladistic classification with cladistic methods of phylogenetic inference.

[7] Figure 2 is from L. L. Cavalli-Sforza et al., *The History and Geography of Human Genes* (Princeton: University Press, 1994), p. 80. For similar models, see Nei and Roychoudhury; Linda Vigilant et al., "African Populations and the Evolution of Human Mitochondrial DNA," *Science*, CCLIII (September 1991): 1503–07; Allan Wilson and Rebecca Cann, "The Recent African Genesis of Humans," *Scientific American*, CCLXVI, 4 (April 1992): 68–73.

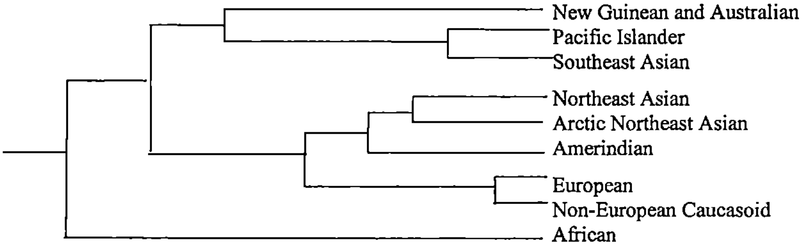


Figure 2: This diagram represents evolutionary relations among human breeding populations.

in part, as reasonably reproductively isolated breeding populations. Yet, because he does not appeal to principles of cladistic classification, there are several important differences between our views. I require that races be monophyletic groups. Consequently, a population must be reproductively isolated over a significant portion of evolutionary history before it can be designated a cladistic race. Kitcher, however, does not require monophyly. He allows the designation of founder populations, so long as (a) the members of such populations show some phenotypic or genetic distinctness, and (b) residual mixed race populations are relatively small. Additionally, I argue that genealogy alone is sufficient for defining 'race', whereas Kitcher holds that it is necessary but not sufficient (conditions (a) and (b) must also be met). Third, Kitcher supports his view using rates of interbreeding among groups in the U.S. today. I, in contrast, use current work in human evolution. Hence, Kitcher is more optimistic about the existence of human races today. Indeed, I hold that races once existed, but may be on their way to becoming extinct.

Joshua Glasgow has recently questioned the viability of our accounts.[8] Although his argument focuses primarily on CRC, his worry is the same for any account that defines races as lineages of breeding populations. Glasgow does not deny that human clades exist, nor does he deny their biological significance. Rather he argues that CRC deviates too far from CS conceptions of race. Glasgow's objection raises important questions about the relation between the scientific and CS meanings of *natural kind* (NK) terms. Who decides the meaning of 'race', or any other NK term, when scientific and folk meanings diverge?[9] In what follows, I address this question and argue that CRC is legitimately a theory about race. It is important to note, however, that much of what I say works in defense of Kitcher's model or any similar conception of race.

---

[8] "On the New Biology of Race," this JOURNAL, C, 9 (September 2003): 456–74.

[9] It is sometimes assumed that genealogical entities are not NKs. See Richard Boyd, "Homeostasis, Species, and Higher Taxa," in Robert Wilson, ed., *Species* (Cambridge: MIT, 1999), pp. 141–85, for an explanation of their compatibility.

I

Glasgow's argument proceeds in two steps. First, he argues that CRC deviates both extensionally and intensionally from CS notions of 'race'. For example, he contends that the number of races recognized by CRC, which he assumes is nine, exceeds the number recognized in CS. While admitting some disagreement within CS over the number of races, Glasgow maintains that rarely does the number exceed five. Second he reminds the reader that CRC raises the possibility that cladistic races cross-classify CS racial groupings.[10] To see why this is so, recall that CRC defines races monophyletically. Though the reconstruction of human evolution is still under way, according to one model, 'Asian' is not a cladistic race. Referring to the phylogeny depicted in figure 2, there is no monophyletic group that includes both Northeast and Southeast Asians that does not also include Caucasoids. In addition to these claimed extensional mismatches, Glasgow argues that there are important intensional mismatches. The main problem, on his view, is that observable traits such as skin color are an inextricable part of the folk race concept. CRC, however, defines 'race' in terms of genealogy alone. Thus, according to Glasgow, CRC treats overt morphology as irrelevant.

Next, Glasgow maintains that because the divergence from CS is extensive, cladistic 'races' are not really races. To make his case, he turns to the following question: "How revisionist can one be about the meaning of 'race' and still call it 'race'" (462)? He answers that minor revisions in meaning are allowable, but when there is too much deviation from CS, the scientific term fails to refer. Glasgow also considers (and rejects) semantic deference to science. He maintains that when deciding the meaning of NK terms, we defer to experts only under certain conditions (which I will discuss below). He adds that CRC cannot meet these conditions and concludes that CS provides the meaning of 'race'. This is just a brief overview of Glasgow's argument. Now let us examine it more closely.

II

Some of Glasgow's objections have been raised before. I will only briefly discuss my earlier responses here.[11] What I argue instead is that Glasgow tells a selective story about the CS meaning of 'race.' Once we examine a fuller picture, we see more overlap between CS and CRC than Glasgow allows. Later I argue that, even in the face of some disagreement, we should call cladistic races 'races'.

---

[10] Andreasen, "New Perspective," and "Race."

[11] Andreasen, "Race," and "Cladistic Race Concept."

Let us begin with Glasgow's "intensional mismatch argument." Recall that his main worry is that CRC treats overt morphology as irrelevant for defining 'race'. Though I agree that overt morphology is often an important part of CS, Glasgow has overlooked other CS definitions that focus solely or primarily on genealogy. The *Oxford English Dictionary*, for example, identifies two dominant, yet distinct, meanings: (A) "A group of persons, animals, or plants, connected by common descent," and (B) "A group or class of persons, animals, or things, having some common features." Moreover according to some historians, these two definitions have been at work in science and CS throughout much of the term's history.[12] For instance, races are defined solely in terms of ancestry in the debate over polygenesis versus monogenesis. Ancestry is also central to the one drop rule, or any other definition that uses blood lines for individuating races. Today the one drop rule has been discredited, but ancestry remains central to some CS notions of race. For example, it is often part of CS that two individuals are members of race *R* if their parents are members of *R*.[13] Finally, Glasgow himself provides examples of folk conceptions that rely on genealogy alone. He cites two distinct legal definitions—one defines 'race' in terms of morphology, the other defines 'race' in terms of descent—and also mentions an argument for the view that, in the U.S., CS conceptions of race are best captured in terms of origins (459–61, notes 11 and 16).[14]

Glasgow needs to explain why, when assessing CRC, we should privilege CS conceptions that rely on morphology over those that rely on ancestry. Glasgow might respond that most genealogical definitions implicitly rely on morphology since they require an independent way to identify the race of one's ancestors (460, note 11). Yet, this type of response will not work. There are other independent criteria, such as geographic origins, for determining the race of one's ancestors. To call someone 'black', for example, is to say that her ancestors come from Africa. Indeed, geographic origins are an important part of CRC. Recall that cladistic races are monophyletic groups of breeding populations, where a breeding population is defined in terms of reasonable reproductive isolation and geographic location. Furthermore contrary to Glasgow's suggestion, overt morphology is not irrelevant to genealogical definitions of race (CS or cladistic). It is often a part of CS that skin color and other overt characteristics are a *heuristic* for identifying racial ancestry. Many Americans presume, for example, that someone with dark skin, has ancestral origins in Africa. Likewise, as I have argued elsewhere, phenotypic and genetic similarity may not *define* cladistic races, but it is used as *evidence* for race membership.[15] Moreover, there is often considerable agreement between classifications based solely on similarity and those based on genealogy alone. Nonetheless when the two disagree, CRC favors the genealogical classification scheme.

---

[12] Banton and Harwood, pp. 13–42; Goldberg.

[13] Kitcher, "Race"; Michael Hardimon, "The Ordinary Concept of Race," this JOURNAL, C, 9 (September 2003): 437–55.

[14] Though Glasgow adds that the groups identified differ from cladistic races because they include some ethnic and national groups, the reader should be aware that, according to many scholars, ethnic and national groups are not racial groups. See, for example, Goldberg; Michael Omi and Howard Winant, *Racial Formation in the United States*, Second Edition (New York: Routledge, 1994).

Not only does Glasgow overlook folk conceptions based solely on ancestry; he also wrongly assumes that the history of intellectual theorizing is irrelevant to understanding the meaning of 'race' (459, note 9). Yet, if we examine this history, we can see that CRC fits squarely within this tradition. The geographical subspecies concept, which defines races as geographically localized phenotypically or genetically distinct breeding populations, was the dominant biological race concept until fairly recently. While this definition has been rejected, several revised definitions have been proposed, all of which define 'race' genealogically.[16] Additionally, many theorists who reject the biological reality of race often accept the legitimacy of genealogical definitions. What they argue instead is that, although nonhuman races (defined genealogically) may exist, no such populations exist in humans.[17]

Because Glasgow brackets the intellectual history of race, the fact that CRC agrees with it might seem irrelevant. Yet Glasgow provides no justification for disregarding this history, and there are good reasons to take it seriously. Understanding a term involves grasping what most competent speakers mean when they use that term. Depending on one's theory of meaning, this can be done descriptively, referentially, or both. Either way, scholars often use intellectual history as an indicator of the CS meaning of 'race.'[18] One reason is that it is often a part of CS that 'race' is a biological kind term. Furthermore, although the CS and scientific histories of 'race' may be somewhat distinct, they are not completely autonomous. When providing biological definitions of 'race', scientists often use CS as a starting point. Likewise, CS conceptions of race have been greatly influenced by science.

---

[15] Andreasen, "Race," and "Cladistic Race Concept."

[16] See note 6.

[17] Lewontin, "Apportionment"; Appiah, pp. 71–74. See Andreasen, "New Perspective," "Race," and "Cladistic Race Concept" for some responses.

[18] See note 1. See also Appiah.

Now let us turn to Glasgow's extensional mismatch argument. My main objection here is that the CS extension of 'race' is not as clearcut as Glasgow makes it seem. He maintains that CS recognizes a minimum of three races (Africans, Asians, and Caucasians) and sometimes recognizes five (by adding Native Americans and Latinos). According to some scholars, however, there is significant confusion within CS over how many races there are and who belongs to what race.[19] Americans have shown uncertainty, for example, over whether certain ethnic or national groups (for example, Mexicans, Latinos, Jews, Filipinos, the Irish) are races. Additionally, the racial categories identified by the U.S. census have changed significantly over time. I have discussed several examples of such changes elsewhere and others are discussed below.[20] Finally, as Glasgow himself acknowledges, the intellectual history of race reveals some uncertainty over the extension of 'race' (459, note 9). In terms of numbers, the lists have varied from as few as three to as many as eighty. There is also significant variability in the racial groups identified. The point of these examples is not to argue for extensional agreement between CRC and CS, at least not yet. My point is that we need an argument as to why we should prefer the classification scheme specified Glasgow over others that have been proposed.

As with his intensional mismatch argument, Glasgow's selective story about the extension of 'race' results in an overstatement of the divergence between CS and CRC. For instance, Glasgow maintains that CRC is committed to the existence of nine races, whereas folk conceptions typically recognize five or less. The accuracy of this claim, however, is called into question once we consider the U.S. census and the intellectual history of race. At different times, the census has recognized between three and eight racial categories. And, as just noted, the intellectual history of 'race' shows significantly more variation. Even if this were not so, I am not committed to there being nine races. What I maintain is that there is a nested hierarchy of cladistic races. Assuming the accuracy of the diagram depicted in figure 2, there are five major races (Africans, Caucasians, NE Asians, SE Asians and Pacific Islanders, and Native Americans) as well as subraces within these groups. Moreover, when we compare the cladistic races listed above to those identified by the 2000 census and directive 15 (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White), we see a good deal of agreement.[21] These sources also recognize subraces similar to those recognized by CRC.

---

[19] Omi and Winant; Smedley; Goldberg; Lawrence Wright, "One Drop of Blood," *The New Yorker* (July 25, 1994): 46–55.

[20] Andreasen, "Cladistic Race Concept."

Now I will respond to Glasgow's worry that cladistic races cross-classify CS racial groupings. Let me begin by reminding the reader that the reconstruction of human evolution is still underway. The branching diagram depicted in figure 2 provides an example of a phylogeny that could be used to identify cladistic races, but CRC does not depend on the accuracy of this tree. In fact, other research groups have provided phylogenies that result in little to no cross-classification.[22] Yet, I continue to maintain that some cross-classification is not a problem for my view. One reason is that differing folk conceptions of 'race' sometimes cross-classify one another as well.[23] When we consider changes in the way the U.S. government has classified by race, we see some examples. Take revisions to the U.S. census, for example. In 1980, Asian Indians were removed from 'Caucasian' and were placed in the 'Asian' category. In 2000, Native Hawaiians and Pacific Islanders were removed from the 'Asian' category and were given independent racial status. These examples illustrate that, as with CRC, there is sometimes disagreement within CS over who counts as Asian. Another example comes from the Equal Employment Opportunity Commission, which treats North Africans as 'white', though many Americans would designate them 'black'. Finally, as I argue below, there are many other examples of cross-classification between the scientific and CS meanings of NK terms.

I take myself to have established reasonable overlap between CRC and CS conceptions of race. Given Glasgow's assertion that minor revisions in meaning are allowable, this is sufficient to show that CRC is legitimately a theory about race. Nonetheless, because I maintain that some divergence is possible, I will now argue that in the face of divergence, we should still call cladistic races 'races'.

---

[21] Directive 15 aims to standardize the collection of racial data among federal agencies.

[22] Andreasen, "New Perspective."

[23] Andreasen, "Cladistic Race Concept."

III

Ultimately, the disagreement between Glasgow and myself is over how meaning gets settled in the first place. When scientists and layfolk disagree, who decides the meaning of NK terms? Who decides the meaning of 'race'? In response to the latter question, Glasgow defends the authority of CS and argues that scientists are not the arbiters of the meaning of 'race'. I, however, have previously argued that we need not reject CRC simply because it deviates somewhat from CS.[24] In this paper, I take a stronger stance. I argue for the relative autonomy of the scientific and folk meanings of 'race'.

Let us begin with my claim that we need not reject CRC because it deviates somewhat from CS. Elsewhere I used the causal theory of reference as well as examples from the history of science to support my point.[25] According to the causal theory, the extension of a NK term is fixed in a baptismal procedure. The kind is then defined by its underlying nature. Since ordinary speakers need not know the underlying nature, they can be competent users of a NK term even if they have mistaken beliefs about the nature of the kind. I also argued that my point does not depend upon the causal theory. The history of science also provides examples of divergence between science and CS over the meaning of NK terms: it is (has been) a part of CS that glass is a solid, whales are fish, bats are birds, species have essences, and the heavenly bodies are immutable. In science, however, glass is a liquid, whales and bats are mammals, species are lineages, and the heavenly bodies are changeable. In such cases, people seldom conclude that the kind does not exist, nor do they request that science be revised to fit CS.

Glasgow rejects my arguments and ultimately argues for privileging CS over science. He begins by claiming that minor revisions in meaning are allowable, but when scientific meaning deviates too far from CS, we are no longer talking about 'race'. Glasgow's intensional and extensional mismatch arguments are meant to show that CRC does indeed deviate too far. He then responds to my history of science analogy by arguing that there are important differences between CRC and the examples listed above. Using 'whale' as an example, he states that the disagreement between scientists and layfolk is over how to classify one anomalous species in an otherwise fixed classification scheme. When it comes to 'race', however, he maintains that CRC involves a "wholesale reshuffling" of the classification scheme itself as well as disagreement over its underlying definition (463).

---

[24] Andreasen, "Race."
[25] Andreasen, "Race."

Second, Glasgow considers what he takes to be the main alternative, namely, the semantic deference to science. He argues that when determining the meaning of NK terms, we give preference to experts over CS only under certain conditions: (A) experts must have an agreed upon definition of a term, and (B) scientific categories must overlap reasonably with CS categories. He adds that CRC fails to meet both conditions. Scientists not only disagree over how to define 'race'; they disagree on whether race is a good taxonomic category in the first place. Moreover, the races identified by the cladistic view fail to pick out reasonably overlapping objects with those identified in CS. Indeed, Glasgow maintains that CRC requires derigidifying 'race'. The races that I identify, which he takes to be the lineages depicted in figure 2, are, on his view, not the ones that were baptized (African, Asian, and Caucasian).

There are a number of problems with Glasgow's argument. First, it relies on his intensional and extensional mismatch arguments which, as I argued, overstate the divergence between CS and CRC. Second, though I am not defending semantic deference to science, there are problems with Glasgow's conditions. Take condition A. Disagreement amongst scientists should not be a problem for semantic deference, provided that there is a fact of the matter about the nature of the kind. And, of course, some experts may know the true nature, even if they fail to agree. Furthermore, reliance on this condition results in a double standard. On Glasgow's view, failure to meet condition A is sufficient for giving priority to CS conceptions of race, but as noted above CS is no different from science when it comes to confusion over the meaning of 'race'.

My objection to the second condition is not with the condition itself, but with Glasgow's application of it. Specifically, because the origins of the race concept are obscure, Glasgow's claim that CRC requires derigidifying 'race' should not be accepted uncritically. Historians disagree, for example, whether 'race' has Arabic, Latin, or German origins; they also disagree over the date of origin.[26] Indeed, most historians hold that 'race' is a hazy concept that has assumed a diversity of meanings (for example, breed or stock of animals; species or kind of animal; tribe; clan; group of people who share a common ancestor; group of people who share common physical features). It is likely that we do not know enough about the history of 'race' to know what was in the minds of speakers during the baptismal procedure.

---

[26] See note 1.

I also object to the conclusion that Glasgow draws from a term's failure to meet the second condition. Glasgow maintains that if the extension of a NK term fails to overlap reasonably with the object(s) identified in CS, we ought to reject the NK term as nonreferring. But as my history of science examples illustrate, sometimes when there is disagreement over the extension of a NK term, people accept divergence between science and CS. Though Glasgow has a response to this point—that my history of science analogy is tenuous—his response is problematic. As already argued, there is no reason to accept his claim that CRC involves a "wholesale reshuffling" of CS racial categories. Additionally, he bases his argument on the category 'whale' when he should be focusing on 'mammal'. Once we correct for this error, the analogy is stronger than Glasgow recognizes. Not only is there disagreement between science and CS over the extension of 'mammal' (bats, dolphins, whales, and platypuses are a few examples of species that layfolk would not normally count as mammals); there is also disagreement over the definition of the term. As with any taxonomic category, most systematists would define 'mammal' in terms of common ancestry. Layfolk, however, typically use similarity as the basis of biological classification schemes.[27] Finally, Glasgow's argument for a disanalogy relies on a single example. Even if he were correct, as I argue below, there are many other examples of extensional disagreement (without revision) between scientific and CS uses of biological kind terms.

Before turning to this point, I would like to raise another objection. Glasgow considers two answers to our question about meaning: when scientists and layfolk disagree over the meaning of NK terms, we can privilege CS over science, or we can privilege science over CS. As we have seen, Glasgow defends the first option; he takes me to be defending the second (463–69). Not only is this a misrepresentation of my view; Glasgow has overlooked a third option. I have not argued for revision to the CS meaning of 'race'. What I argued is that 'race' is ambiguous between its scientific and CS meanings (with reasonable overlap between the two).[28] Implicit, though not directly argued for, was the idea that these meanings are relatively autonomous. I will develop this argument momentarily. For now let me stress that by overlooking this option, Glasgow's argument does not apply to my view.

---

[27] Scott Atran, *Cognitive Foundations of Natural History* (New York: Cambridge, 1990).
[28] Andreasen, "Race," pp. S661–65.

Glasgow's misrepresentation of my view might be due to my reliance on the causal theory of reference, so let me clarify the role that it is supposed to play. To do so, we must take a closer look at the ambiguity of 'race'. Elsewhere, I argued that the term currently functions differently in science and CS. In science, race is a taxonomic category that helps to explain patterns of migration, reaction to adaptive pressures, and the history of human evolution.[29] In CS, race helps to explain human social relations, racist beliefs and practices, and the like. Because its explanatory role is different in each case, 'race' has come to function as a NK term in one context and a social kind term in another. Its meaning as a NK term is given by science; its meaning as a social kind term is given by CS. Since there is significant overlap between these uses, both deserve the label 'race'. Returning to the causal theory, it tells us that when scientific and folk meanings of NK terms diverge, we need not revise science to fit CS. What it says about social kind terms is an open question.

Now I will develop my argument for the relative autonomy of scientific and CS conceptions of 'race'. As John Dupré and Scott Atran have argued, divergence between folk biological and systematic categories is not uncommon.[30] Take the term 'lily', for example. In science, 'lily' refers to numerous genera that make up the lily family, including many species that would not normally be recognized as lilies (for example, some tulips, onions, and garlics). Dupré and Atran offer many other examples of this sort. They add that in such cases, there is often no revision (nor any expectation of revision) to science or CS. This alone provides reason to accept divergence, without revision, between the CS and scientific meanings of 'race'.

A further argument can be advanced as well. According to Atran and Dupré, divergence among systematic and folk biological categories occurs for good reasons. To see why this is so, let us consider the history of systematic biology. Within systematics, taxa were originally defined in terms of similarity; examples include essentialism and pheneticism. This practice was eventually superseded by phylogenetic classification, partly because lineages play an important role in evolutionary theorizing. Ordinary language, however, did not follow suit. Many folk biological categories are based solely on observable properties, partly because they are useful for explaining the readily observable world.[31] Since classification schemes based on similarity sometimes disagree with those based on ancestry, the result is some divergence between science and CS. Moreover, since folk biological and systematic classification schemes serve different functions (explaining the observable world versus explaining evolutionary phenomena), there is no reason to expect convergence. This brief history is important because it parallels what I have said about race.[32] Biological races were originally defined in terms of similarity. Examples include the typological and geographical race concepts. These definitions have been rejected, but there is no reason to stop short of a genealogical definition of 'race'. Also, since these conceptions serve distinct functions, they can coexist in relative harmony.

---

[29] See note 6.

[30] See Atran; Dupré, *The Disorder of Things* (Cambridge: Harvard, 1993).

IV

Glasgow asks for an argument to show that cladistic races are more like whales (or mammals) than like witchcraft. I take the above considerations to meet this challenge. Not only is there reasonable overlap between CRC and CS conceptions of race; some divergence is not a reason to deny that cladistic races exist. Cladistic races are indeed more real than witchcraft.

ROBIN O. ANDREASEN

University of Delaware

---

[31] See Atran.

[32] See Andreasen, "New Perspective," and "Race."

NOTES AND NEWS

The Editors note with sorrow the death of Jacques Derrida on October 8, 2004, at the age of 74. Born in El-Bair, Algeria, Derrida graduated in 1956 from the École Normal Supérieure in Paris, coming to America briefly as a visiting scholar at Harvard University. He returned to Paris to teach at the Sorbonne in 1960, joining the faculty of École Normal Supérieure in 1965 where he remained until 1984. Beginning in the 1970s, he began dividing his time between the École Normal Supérieure and regular visiting appointments in the U.S., first at Johns Hopkins University, then at Yale University from 1975–1986. He went on to take an appointment at the University of California, Irvine, which he held until 2003, when illness required he reduce his teaching schedule.

Derrida was a remarkably prolific writer, publishing countless essays and well over 40 books in French. He first came to the attention of American academia with his 1966 lecture at Johns Hopkins entitled "Structure, Sign, and Play in the Discourse of the Human Sciences." The following year, he published three major works: *Speech and Phenomena, Of Grammatology,* and *Writing and Difference.* Other major publications include *Positions, Dissemination, Margins of Philosophy* (1972), *Glas* (1974), *Spurs: Nietzsche's Styles* (1979), *The Postcard: From Socrates to Freud and beyond* (1980), and *The Truth in Painting* (1987).

Derrida was very active in the intellectual life on the Continent, and was involved in work against apartheid in South Africa and supporting the rights of Czech dissidents. He was one of the conveners of the Estates General of Philosophy in 1979, and helped found the Collège International de Philosophie in 1983, becoming its first director. Also in 1983, Derrida was elected directeur d'études at the Ecole des Hautes Études en Sciences Sociales in Paris, a position he held until his official retirement at the age of 67. Derrida was elected to the American Academy of Arts and Sciences as an Honorary Foreign Member and was awarded honorary degrees from numerous institutions.

The Editors note with sorrow the death of Paul Edwards in early December 2004, at the age of 81. He received his B.A. and M.A. from the University of Melbourne and received his Ph.D. in philosophy from Columbia University in 1954. Edwards taught at New York University in the 1960s, at Brooklyn College from 1966 to 1986 and at the New School for Social Research from the late 1960s into the 1990s. Edwards was the author of *The Logic of Moral Discourse* (1955), *Buber and Buberism: A Critical Evaluation* (1971), *Reincarnation: A Critical Examination* (1996), and *Heidegger's Confusions* (2004). He is best known as the editor of *The Encyclopedia of Philosophy,* a multi-volume work first published in 1967 and still very much in use to this day.

NEW BOOKS

TANESINI, ALESSANDRA: *Wittgenstein: A Feminist Interpretation.* Malden, MA: Polity, 2004. xiv, 164 p. Paper $29.95.

TEMES, PETER S.: *The Just War: An American Reflection on the Morality of War in Our Time.* Chicago: Ivan R. Dee, 2004. xv, 224 p. Paper $14.95.

THOMPSON, DENNIS F.: *Restoring Responsibility: Ethics in Government, Business, and Healthcare.* New York: Cambridge, 2004. viii, 349 p. Cloth $70.00, paper $25.99.

TROPPER, AMRAM: *Wisdom, Politics, and Historiography: Tractate Avot in the Context of the Graeco-Roman Near East.* New York: Oxford, 2004. xi, 302 p. Cloth $99.00.

TUCKER, AVIEZER: *Our Knowledge of the Past: A Philosophy of Historiography.* New York: Cambridge, 2004. 291 p. Cloth $70.00.

TURNER, DENYS: *Faith, Reason and the Existence of God.* New York: Cambridge, 2004. xix, 271 p. Cloth $70.00, paper $27.99.

VISION, GERALD: *Veritas: The Correspondence Theory and Its Critics.* Cambridge: MIT, 2004. xiv, 298 p. Cloth $36.00.

VOEGELIN, ERIC: *The Drama of Humanity and Other Miscellaneous Papers, 1939–1985.* WILLIAM PETROPULOS and GILBERT WEISS, eds. Columbia: Missouri UP, 2004. vii, 484 p. Paper $59.95.

WEBER, MARCEL: *Philosophy of Experimental Biology.* New York: Cambridge, 2005. xvi, 358 p. Cloth $75.00.

WILSON, ELIZABETH A.: *Psychosomatic: Feminism and the Neurological Body.* Durham, NC: Duke, 2004. x, 126 p. Cloth $64.95, paper $18.95.

WILSON, ROBERT A.: *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition.* New York: Cambridge, 2004. xvii, 369 p. Cloth $70.00, paper $24.99.

WINANT, HOWARD: *The New Politics of Race.* Minneapolis: Minnesota UP, 2004. xx, 275 p. Cloth $59.95, paper $19.95.

WOOD, ALLEN W.: *Kant.* Malden, MA: Blackwell, 2005. xiv, 195 p. Paper $19.95.

YAFFE, GIDEON: *Manifest Activity: Thomas Reid's Theory of Action.* New York: Oxford, 2004. viii, 167 p. Cloth $39.95.

YOLTON, JOHN W.: *The Two Intellectual Worlds of John Locke: Man, Person, and Spirits in the "Essay."* Ithaca: Cornell, 2004. ix, 180 p. Cloth $35.00.

ZIPORYN, BROOK: *Being and Ambiguity: Philosophical Experiments with Tiantai Buddhism.* Chicago: Open Court, 2004. xxii, 452 p. Paper $32.95.

ZUIDERVAART, LAMBERT: *Artistic Truth: Aesthetics, Discourse, and Imaginative Disclosure.* New York: Cambridge, 2004. xiii, 277 p. Cloth $75.00.

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2005

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org**

crítica

REVISTA HISPANOAMERICANA DE FILOSOFÍA

Vol. 36 • No. 107 • agosto 2004 • ISSN 0011-1503

Artículos

HUMBERTO SCHETTINO
The Notion of Politics in Marx's Early Writings

HUGO OMAR SELEME
¿Puede el utilitarismo ser deontológico? Una respuesta a Kymlicka

Discusiones

CARLOS PEREDA
¿Puede la moral prescindir de principios universales? Una discusión con Josep Corbí

JOSEP CORBÍ
En respuesta al comentario de Carlos Pereda sobre Un lugar para la moral

Notas bibliográficas

MARIO GÓMEZ-TORRENTE, *Forma y modalidad. Una introducción al concepto de consecuencia lógica*
[Axel Barceló Aspeitia]

CRÍTICA, Revista Hispanoamericana de Filosofía aparece tres veces al año en los meses de abril, agosto y diciembre. Es una publicación del Instituto de Investigaciones Filosóficas (UNAM). Todo tipo de correspondencia deberá ser dirigida al Apartado Postal 70-447, Coyoacán, 04510, México, D.F. Correo electrónico: critica@filosoficas.unam.mx

CAMBRIDGE

BEST IN OUTSTANDING SCHOLARSHIP

Now in Paperback...

Faking It

William Ian Miller

Praise for the Hardback edition:

"William Ian Miller... scratches the itch of authenticity and relieves the ache of moralit with delicious determination in Faking It" —The Boston Globe

$18.99*: Paperback:
0-521-61370-1: c. 304pp.

Autonomy and the Challenges to Liberalism

New Essays

John Christman, Pennsylvania State University, State College, PA
Joel Anderson, University of Utrecht, The Netherlands
Editors

$75.00*: Hardback:
0-521-83951-3: c. 368pp

The Cambridge Companion to Anselm

Cambridge Companions to Philosophy

Brian Davies, Fordham University, New York, NY
Brian Leftow, University of Oxford, UK
Editors

$65.00*: Hardback:
0-521-80746-8: c. 399pp.
$29.99*: Paperback:
0-521-00205-2

The Cambridge Companion to Arabic Philosophy

Cambridge Companions to Philosophy

Peter Adamson, University of London, UK
Richard C. Taylor, Marquette University, Milwaukee, WI
Editors

$70.00*: Hardback:
0-521-81743-9: c 460 pp.
$29.99*: Paperback:
0-521-52069-X

The Cambridge Companion to Montaigne

Ullrich Langer, University of Wisconsin, Madison, WI
Editor

Cambridge Companions to Philosophy

$65.00*: Hardback:
0-521-81953-9: c. 288pp.
$22.99* Paperback.
0-521-52556-X

How the Cold War Transformed Philosophy of Science

To the Icy Slopes of Logic

George A. Reisch

$70.00*: Hardback:
0-521-83797-9. c. 430 pp.
$23.99*: Paperback:
0-521-54689-3

Medieval Islamic Philosophical Writings

Muhammad Ali Khalidi, American University of Beirut, Lebanon
Editor

Cambridge Texts in the History of Philosophy

$65.00* Hardback:
0-521-82243-2: c. 254 pp.
$25.99*: Paperback:
0-521-52963-8

The Selfish Meme

A Critical Reassessment

Kate Distin

$60.00: Hardback:
0-521-84452-5: 240 pp
$21.99: Paperback:
0-521-60627-6

Evidence and Faith

Philosophy and Religion since the Seventeenth Century

Charles Taliaferro

The Evolution of Modern Philosophy

$75.00 Hardback:
0-521-79027-1: 472 pp.
$29 99: Paperback:
0-521-79375-0

Foucault and Classical Antiquity

Power, Ethics and Knowledge

Wolfgang Detel

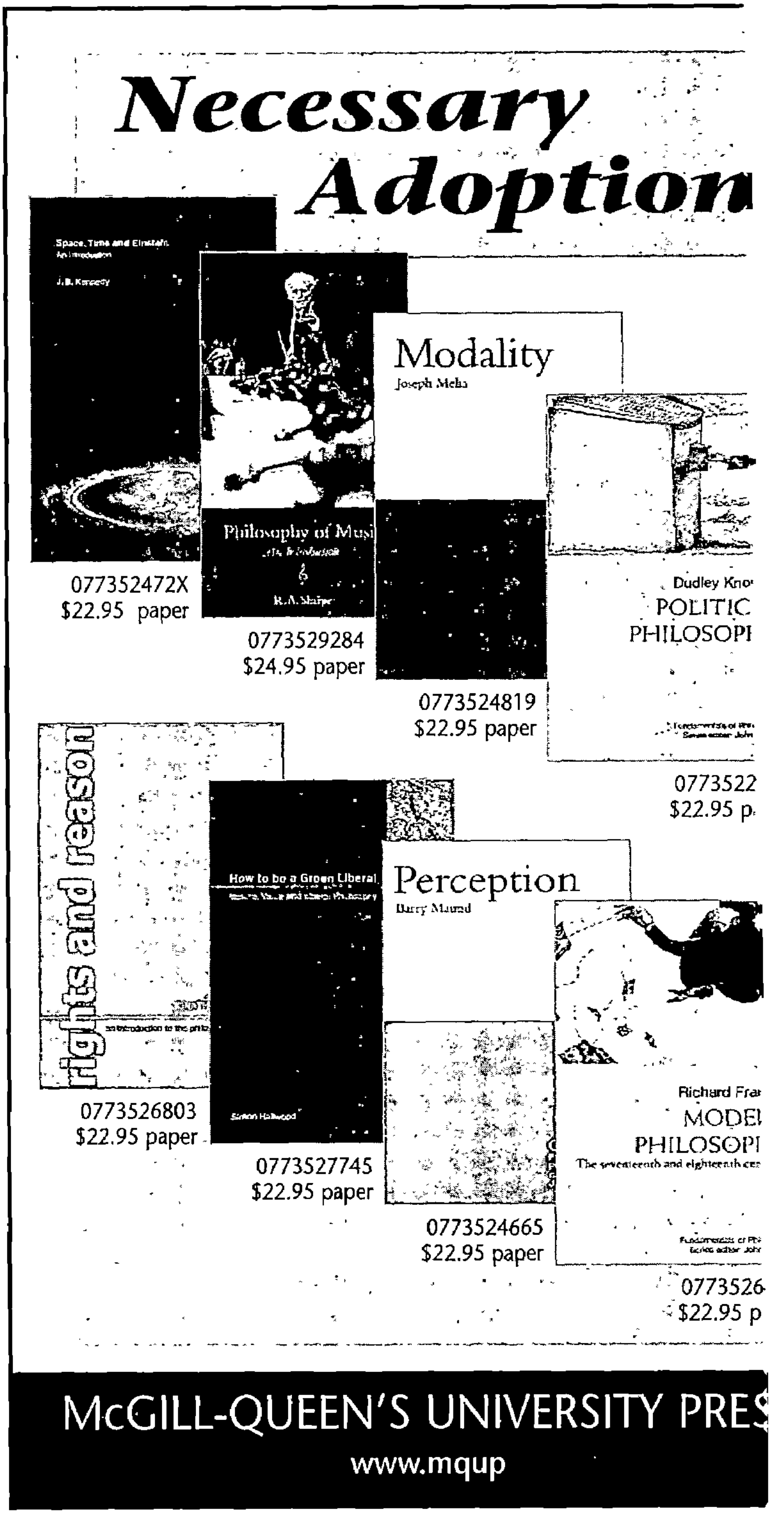
Modern European Philosophy

$75.00. Hardback:
0-521-83381-7: 291 pp

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

Necessary Adoption



0773524 72X
$22.95 paper

0773529284
$24.95 paper

0773524819
$22.95 paper

0773522
$22.95 p

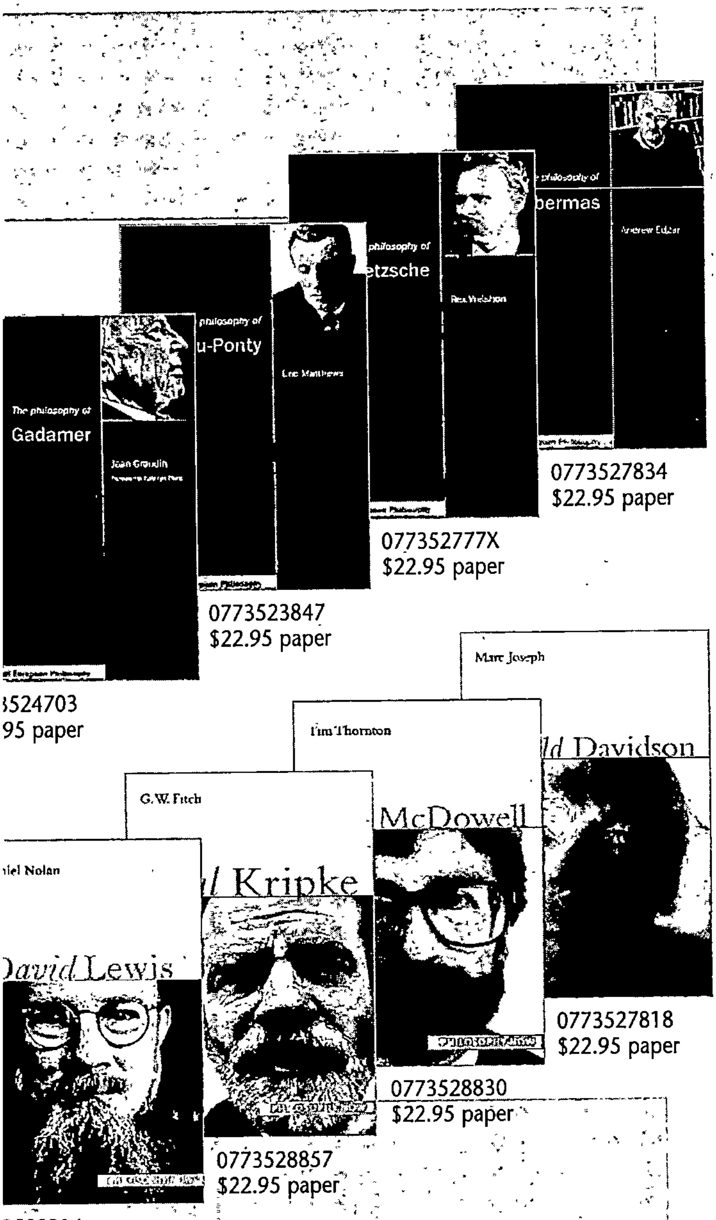
0773526803
$22.95 paper

0773527745
$22.95 paper

0773524665
$22.95 paper

0773526
$22.95 p

McGILL-QUEEN'S UNIVERSITY PRES
www.mqup



3524703
95 paper

0773523847
$22.95 paper

077352777X
$22.95 paper

0773527834
$22.95 paper

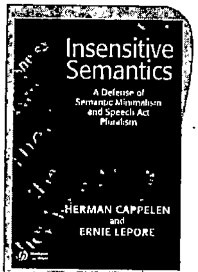
3529314
.95 paper

0773528857
$22.95 paper

0773528830
$22.95 paper

0773527818
$22.95 paper

GILL-QUEEN'S UNIVERSITY PRESS
Distribution by CUP Services 1 800 666 2211


Insensitive Semantics
A Defense of Semantic Minimalism and Speech Act Pluralism
HERMAN CAPPELEN and ERNIE LEPORE

Insensitive Semantics

A Defense of Semantic Minimalism and Speech Act Pluralism

By **HERMAN CAPPELEN** *and* **ERNIE LEPORE**
Vassar College, University of Oslo; Rutgers University

"This book is an ingenious defense of two positions not widely thought to be compatible: truth-conditional semantics and semantic minimalism. Cappelen and Lepore's highly controversial views are already, and will continue to be, at the center of inquiry into the nature of linguistic communication."
~ Jason Stanley, University of Michigan

January 2005
6 x 9 ~ 232 pages
1-4051-2675-2
Paperback ~ $29.99
1-4051-2674-4
Hardcover ~ $64.95

Insensitive Semantics is an overview of and contribution to the debates about how to accommodate context sensitivity within a theory of human communication, investigating the effects of context on communicative interaction and, as a corollary, what a context of utterance is and what it is to be in one.

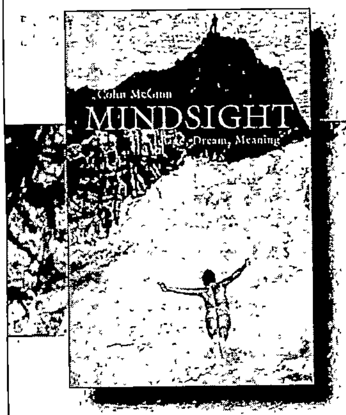
Blackwell Publishing

ORDER TODAY! To purchase a copy of this book, or to request an examination copy, visit the website at: **www.blackwellpublishing.com**

THE JOURNAL OF PHILOSOPHY

The Journal is pleased to announce to our readers and subscribers the launch of its new website, which will have the pages listed below. The address is: *www.journalofphilosophy.org* or simply *www.jphil.org*.

- Purpose & History
- Contacts and Addresses
- Board of Trustees
- Editorial Board
- Memos to Authors
- Subscriptions & Advertising
- Recent & Forthcoming Issues
- Electronic Issues
- Index 1904–present
- Rights and Permissions
- Lectures & Conferences
- Philosophy Links



MINDSIGHT

Image, Dream, Meaning

COLIN MCGINN

Clearly, seeing an object is similar in certain respects to forming a mental image of it, but it is also different. McGinn shows what the differences are, arguing that imagination is a sui generis mental faculty. He goes on to discuss the nature of dreaming and madness, contending that these are primarily imaginative phenomena.

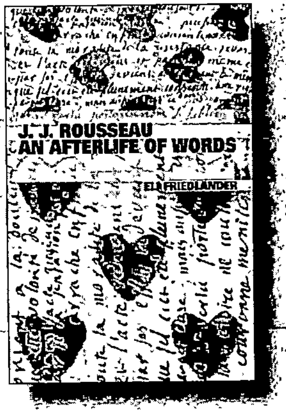
new in cloth

LAW'S QUANDARY

STEVEN D. SMITH

Steven Smith argues that our legal vocabulary and methods of reasoning presuppose classical ontological commitments that were explicitly articulated by thinkers from Aquinas to Coke to Blackstone, and even by Joseph Story. But these commitments are out of sync with the world view that prevails today in academic and professional thinking. So our law-talk thus degenerates into "just words"—or a kind of nonsense.

new in cloth



J. J. ROUSSEAU: AN AFTERLIFE OF WORDS

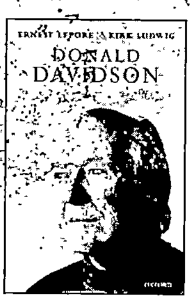
ELI FRIEDLANDER

Eli Friedlander reads Rousseau's autobiography, *Reveries of the Solitary Walker*, as philosophy. Reading this work against Descartes's *Meditations*, Friedlander shows how Rousseau's memorable transformation of experience through writing opens up the possibility of affirming even the most dejected state of being and allows the emergence of the innocence of nature out of the ruins of all social attachments.

new in cloth

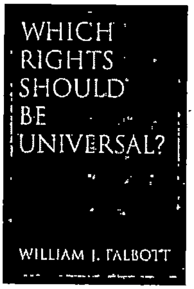
New from
Oxford

New in Paperback!
The Atrocity Paradigm
A Theory of Evil
CLAUDIA CARD
"Ordinarily, I would advise taking the puffery on a book's dust jacket with a grain of salt. But, in this case, I think no exaggeration at all is involved in the claim that the book contains a good deal of wisdom. In sum, this is an excellent book. I recommend it very highly to any philosopher who is interested in the topic of evil."
—*Notre Dame Philosophical Reviews*
paper February 2005 $25.00



Donald Davidson
Meaning, Truth, Language, and Reality
ERNEST LEPORE and KIRK LUDWIG
Davidson's ideas had a deep and broad influence in the central areas of philosophy; he presented them in brilliant essays over four decades, but never set out explicitly the overarching scheme in which they all have their place. Lepore's and Ludwig's book will therefore be the key work, besides Davidson's own, for understanding one of the greatest philosophers of the twentieth century.
February 2005 $60 00

Semantics versus Pragmatics
Edited by ZOLTAN GENDLER SZABO
Leading scholars in the philosophy of language and theoretical linguistics present brand-new papers on a major topic at the intersection of the two fields, the distinction between semantics and pragmatics.
February 2005 paper $29 95 cloth $95 00



Which Rights Should Be Universal?
WILLIAM TALBOTT
In this book, William Talbott builds on the work of John Rawls, Jürgen Habermas, J.S. Mill, Amartya Sen, and Henry Shue to explain how, over the course of history, human beings have learned how to adopt a distinctively moral point of view from which it is possible to make universal, though not infallible, judgments of right and wrong.
February 2005 $35.00

New in Paperback!
Stability and Justification in Hume's Treatise
LOUIS E. LOEB
In his *Treatise*, Hume confronted the tensions between his project of uncovering the causal operations of the human mind and the extreme skeptical tendencies of his system. Louis Loeb argues that Hume overreaches, and he advances a controversial interpretation of Hume's epistemological framework that shows how Hume could have avoided the more destructive positions in his work.
paper February 2005 $25.00

From a Deflationary Point of View
PAUL HORWICH
The papers collected in this volume represent Paul Horwich's development of the deflationary perspective over two and a half decades, and demonstrate its considerable power and fertility.
February 2005 paper $24.95 cloth $60.00
JKC 120104

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

Calcutta ... ALIPORE CAMPUS 1183 21.07.05

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 3
MARCH 2005

page
109 *A Plea for Asymmetric Games* Zachary Ernst
126 *On the Possibility of Paretian Egalitarianism* Bertil Tungodden and Peter Vallentyne

BOOK REVIEWS
155 Christopher Peacocke: *The Realm of Reason* Neil Tennant

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc.

ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 3, MARCH 2005

Calcutta University Library
ALIPORE CAMPUS

A PLEA FOR ASYMMETRIC GAMES*

Game theory has been an important tool for addressing difficult problems regarding the origins of conventions, fairness, and prosocial behaviors in general. Examples of the successful use of game theory include David Lewis's seminal work on conventions,[1] John Nash's theory of equilibrium behavior in noncooperative games,[2] and Robert Axelrod's use of the iterated Prisoner's Dilemma to model the evolution of cooperation.[3]

One feature of game-theoretic models that makes them so useful is their simplicity. However, simple formal models of complex phenomena are a double-edged sword. On the one hand, when an explanation is convincing, and that explanation uses only a simple model of the explanandum, it may possess several important explanatory virtues. For instance, its simplicity may indicate that a variety of features of the phenomenon may be irrelevant or at least unimportant. If so, then we may begin justifiably to suspect that the explanation may apply to other domains. As an example, the same formal tools have been applied successfully to both social learning processes and biological evolutionary processes.

But on the other hand, simple formal models omit details that may turn out to be important. If the devil is in the details, then simple formal models will not provide the robust explanations that we seek. Accordingly, when a simple model has been successfully applied to

---

* Portions of this paper were presented at Texas Tech University, Florida State University, and the California Institute of Technology. I would like to thank the faculty of those departments for valuable discussion. Thanks are also due to Jim Joyce, Sara Rachel Chant, and Michael Ruse.

[1] *Convention: A Philosophical Study* (Cambridge: Harvard, 1969).

[2] "The Bargaining Problem," *Econometrica*, XVIII (1950): 155–62.

[3] *The Evolution of Cooperation* (New York: Basic Books, 1984).

0022-362X/05/0203/109–25

one domain, there is the temptation to apply that technique to another domain for which it may not be appropriate.

Noncooperative game theory is such a simple and powerful tool that this temptation is particularly strong. The basic tenets of game theory can be stated using nothing more than high-school algebra, and even evolutionary game theory can be explained using mathematics no more complex than that found in a first-year calculus course. In spite of this simplicity, it has been deployed to analyze evolutionarily evolved behaviors, social learning processes, the formation of conventions, and the behavior of economic markets, to name just a few of its applications.

It is just a short step from these applications to the task of explaining the origins of our moral practices. After all, our moral practices resemble conventions, and the social contract may be thought of as a set of self-enforcing equilibrium behaviors. Thus, when we observe a practice of following principles such as "share and share alike," a tendency to cooperate in Prisoner's Dilemmas, or to follow broadly utilitarian principles, we may ask whether noncooperative game theory might explain these tendencies.

In this paper, I explore the use of game theory in explaining the origins of our moral practices. Negatively, I will argue that the great generality of game-theoretic explanations has encouraged philosophers to misconstrue the project in such a way as to block progress and invite a raft of criticisms. But positively, I will argue that such excesses are not a necessary component of the project. Specifically, I will argue that the generality of game-theoretic explanations may be abandoned in favor of a legitimate pluralism. That is, game theory may be used to construct a variety of specific explanations of particular moral practices, rather than a single explanation of diverse moral practices.

I. GAMES AND BLACK BOXES

Individuals often cooperate in one-shot Prisoner's Dilemmas, share windfalls even when they are not forced to do so, form coalitions, and follow conventions even when there is no penalty for deviation. Such behaviors require explanation, because the standard *Homo economicus* conception of human beings as perfectly rational, self-interested welfare maximizers predicts that such prosocial behaviors would not be observed nearly as often as they actually are.

There is a simple procedure for using game theory as the primary explanatory tool for explaining the origins of prosocial behaviors. First, we strip away the irrelevant details of the situation, and focus only on the possible actions of the agents, as well as the payoffs the agents will obtain, contingent upon their actions. For instance, experimental economists have performed experiments in which individuals must determine how to split a $10 windfall between them. Not surprisingly, they overwhelmingly favor a 50/50 split, even when the participants cannot communicate with each other. The interesting question here is not "What do the agents do?" for it is obvious that they will split it evenly. Rather, the interesting question is "Why is it so obvious that they will favor the 50/50 split?" In other words, we want to know if there is anything about a 50/50 split that accounts for why it is such an intuitively compelling and expected outcome.

A game theorist such as Brian Skyrms (who has examined this game in detail[4]) will begin by simplifying the situation. From a game-theoretic perspective, it is not important that the subjects are participating in an experiment, that they are splitting money rather than some other divisible good, what their environment is like when they are negotiating, and so on. The only important questions that we consider are "What are the possible actions available to each participant?" and "What are the possible payoffs that accompany those actions?"

Those two questions lead directly to the game-theoretic model. We may suppose (following Skyrms) that each participant may demand either one-third, one-half, or two-thirds of the money. We may settle the second question by supposing that if their demands add up to 100% or less of the money, then each participant gets what she demanded, but if their demands are greater than 100%, then each gets nothing.

Not surprisingly, there are a variety of considerations suggesting that those two facts by themselves are sufficient to favor the 50/50 split. Although those arguments have been questioned in other contexts,[5] I will delay considering those complications until later. What is important at this point is that this game-theoretic explanation omits the mechanisms that are supposed to give rise to the "fair" behavior. For instance, it is possible that the participants' behaviors are the result of social learning processes—perhaps each participant has had the opportunity to observe others in similar situations, and is following the behavior that has led to the best outcomes in those cases. Or perhaps each is hypothetically considering the various alternatives and has hit upon a convincing argument in favor of the 50/50 split.[6] On the other end of the spectrum, perhaps the participants' behavior is the result of a hard-wired behavioral tendency that is the result of evolution and natural selection.

---

[4] "Sex and Justice," this JOURNAL, XCI, 6 (June 1994): 305–20; *Evolution of the Social Contract* (New York: Cambridge, 1996).

[5] Justin D'Arms, "Sex, Fairness, and the Theory of Games," this JOURNAL, XCVI, 12 (December 1996): 615–27; D'Arms, Robert Batterman, and Krzyzstof Gorny, "Game Theoretic Explanations and the Evolution of Justice," *Philosophy of Science*, LXV (1998): 76–102; Martin Barrett, Ellery Eells, Branden Fitelson, and Elliott Sober, "Models and Reality: A Review of Brian Skyrms's *Evolution of the Social Contract*," *Philosophy and Phenomenological Research*, LIX (1999): 237–42.

As it stands, the game-theoretic explanation treats all of these mechanisms as black boxes. Instead of focusing on the particular mechanisms that may be the proximate causes of the observed behavior, the game theorist assumes as a working hypothesis that there is something about the available strategies and payoffs themselves that is sufficient to explain and predict the observed behavior. If so, then any causal process that operates in such a situation will favor the 50/50 split, even if we do not specify whether the mechanism is social or biological. Accordingly, we have Skyrms's representative comment that social and biological processes may be treated analogously:

> These biological concepts also have qualitative analogues in the realm of cultural evolution. Mutation corresponds to spontaneous trial of new behaviors. Recombination of complex thoughts and strategies is a source of novelty in culture. Using these tools of evolutionary dynamics, we can now study aspects of the social contract from a new perspective.[7]

If Skyrms is correct, then game-theoretic explanations of prosocial behaviors may be extremely powerful. Although it tempting to charge that Skyrms has simply disregarded the relevant differences between biological and social processes, this objection oversimplifies his position. According to game theorists, the real explanatory burden is to be placed upon the payoffs and stability properties of the interaction itself, and not on the proximate causes of behavior. This is not an unmotivated position, for the ability to subsume diverse phenomena under a single explanatory argument pattern is often taken to be the hallmark of explanation.[8]

Furthermore, there is another important reason why game theorists disregard the specific mechanisms that may play a role in the evolution of prosocial behaviors. Specifically, the mechanisms that typically accompany game-theoretic explanation tend to converge on the same predictions for a variety of games. And as a general point, if a variety of proximate mechanisms can all be shown to lead to the same observed phenomena, then one can argue that those mechanisms may be omit-

---

[6] As in Nash's original conception of the bargaining process.

[7] Skyrms, *Evolution of the Social Contract*, p. x.

[8] See Philip Kitcher, "Explanatory Unification," in Joseph Pitt, ed., *Theories of Explanation* (New York: Oxford 1988), pp. 167–87.

	Stag	Hare
Stag	7,7	0,3
Hare	3,0	2,2

Figure 1: The Stag Hunt. Payoffs to the row player are first; payoffs to the column player are second.

ted from explanations.[9] Indeed, this seems to be the case regarding the three types of proximate mechanism that are usually invoked in evolutionary explanations of prosocial behavior: equilibrium, assortation, and efficiency. We may illustrate each in turn.

I.1. Equilibrium. The type of explanation that was pioneered by Nash (*op. cit.*) in the 1950s does not advert to the causal history of a population of individuals, and it deliberately sets aside the issue of how the individuals in a population are supposed to converge on a prosocial behavior. Rather, equilibrium explanations make use of the fact that some collective behaviors are more resistant to change than others.

For example, consider the Stag Hunt game. This game is used to model a situation in which two players can either cooperate to achieve a high payoff, or go on their own to achieve a low, but guaranteed payoff. The game is represented in Figure 1, and is traditionally accompanied by the following story. There are two hunters, who each face a choice between hunting a stag and hunting a hare. It takes two hunters working together to hunt successfully a stag, so they will only achieve the higher payoff if both cooperate by hunting stag. However, each may strike out on her own and hunt a hare, for a low, but guaranteed, payoff.

The Stag Hunt provides a simple illustration of equilibrium explanations in game theory. Consider a population in which half the individuals play Stag, and the other half play Hare. Furthermore, suppose that individuals pair up randomly, play the game with each other once, and then disperse back into the population. The Stag players will receive a payoff of 7 half the time, and a payoff of 0 the other half, for an expected payoff of 3 1/2. Hare players will receive a payoff of 3 half of the time and a payoff of 2 the other half of the time. Thus, Hare players will receive an expected payoff of 3/2, which is less than the Stag payoff.

---

[9] Compare with Sober's notion of an "equilibrium explanation" in "Equilibrium Explanation," *Philosophical Studies*, XLIII (1983): 201–10.

Thus, under reasonable assumptions, we can expect this state of affairs to be unstable. If we are working within a social learning context and the players seek to maximize their own payoffs, then those who play the Hare strategy have an incentive to switch strategies. If we are thinking in terms of biological evolution and natural selection, then Stag players will have a higher average fitness than Hare players. And under reasonable assumptions, we would expect the population to approach a state in which everyone plays Stag.

On the other hand, a state in which everyone plays Stag is a stable equilibrium. In such a population, each player gets a payoff of 7 in every interaction. Thus, if someone were to switch to a strategy of playing Hare, her payoff would go down from 7 to 3. Because no individual has an incentive to switch strategies in a population where everyone plays Stag, we say that the Stag strategy is a Nash equilibrium.

Similarly, the state in which everyone plays Hare is also a Nash equilibrium. For in such a population, anyone who switches from Hare to Stag will lower her payoff from 2 to 0. However, here the notion of a stochastically stable equilibrium is useful.[10] If the population is subject to random shocks, it is easier for the population to be moved from the Hare equilibrium to Stag equilibrium than vice-versa. Thus, we say that the Stag equilibrium is the unique stochastically stable Nash equilibrium, and is therefore the one we should expect the population to spend most of the time in over the long run.

I.2. Assortation. Along with stability properties, mechanisms of assortation have justifiably taken center stage in game-theoretic explanations of prosocial behavior. An assortation mechanism is any mechanism that raises the probability that similarly behaving individuals will interact with each other. These mechanisms may take the form of a tendency to interact with genetic relatives with similar genetically programmed behavioral tendencies, or a tendency deliberately and rationally to seek out like-minded individuals in a social context.

If there is such an assortative mechanism, then the expected payoff of the Stag strategy will be significantly higher than the expected payoff of the Hare strategy. For when two Stag players interact, each receives a payoff of 7, whereas when two Hare players interact, each receives a payoff of only 2. So if the assortation mechanism operates to a sufficiently high degree, the Stag strategy is the one we would expect to observe over the long run.

---

[10] For a clear exposition of the concept of a stochastic stability, see H. Peyton Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton: University Press, 1998). See also Larry Samuelson, *Evolutionary Games and Equilibrium Selection* (Cambridge: MIT, 1997).

As is well appreciated, assortation favors the intuitively "fair" or "just" strategy in a variety of different games. For example, it is one of the few plausible mechanisms that cause cooperation to evolve in the one-shot Prisoner's Dilemma. For the cooperative strategy in the Prisoner's Dilemma is the one that maximizes the overall payoff when it is employed by all the players. In contrast, two players who both use the uncooperative strategy will achieve only very low payoffs. Interestingly, this seems to be a general feature of the games that model prosocial behaviors. That is, the prosocial behaviors are normally the ones that maximize payoffs, under the assumption that everyone plays the same strategy.[11]

I.3. Efficiency. A more controversial mechanism, usually considered only in a biological context, is group selection. Although there is a good deal of controversy regarding the nature of group selection mechanisms—and indeed whether there are group selection mechanisms at all—we can give them a general characterization that is adequate for our purposes here. Group selection mechanisms are said to operate when there is some environmental pressure that forces one group of conspecifics to compete with each other for survival and reproduction. As Elliott Sober and David Sloan Wilson have put it, group selection operates whenever a group of individuals are "in the same boat" as regards survival and reproduction.[12]

It is a difficult and complicated matter to model intergroup competition. But many such models have converged on the use of efficiency as a measure of what we might call "group fitness." Efficiency, as understood by economists, is simply a measure of how little of a resource is wasted in interactions. When a group of individuals plays a set of strategies that wastes very little, then we say that the group has high efficiency. In a game-theoretic context, efficiency is measured simply by taking the sum of the payoffs in an interaction.

Indeed, this interpretation of the concept of efficiency makes sense if payoffs are interpreted as fitness. So if a group has high average payoffs, then the group will tend to reproduce more quickly. And higher rates of reproduction make it less likely that a group will face extinction, be outcompeted for scarce resources, and so on.

Of course, it is apparent that efficiency considerations will favor prosocial behaviors like that represented in the Stag Hunt game. For the highest possible efficiency is obtained to the degree to which a population tends to play the Stag strategy. Similarly, the intuitively "fair," "just," "cooperative," or "altruistic" behaviors in other games such as the Prisoner's Dilemma are also favored by efficiency considerations. We may conclude that if group selection mechanisms operate to a sufficiently high degree, and they favor efficient strategies, then we should expect the prosocial behaviors to evolve over time.

---

[11] Skyrms, "Darwin Meets *The Logic of Decision*: Correlation in Evolutionary Game Theory," *Philosophy of Science*, LXI (1994): 503–28.

[12] "Reintroducing Group Selection to the Human Behavioral Sciences," *Behavioral and Brain Sciences*, XVII (1994): 585–654. But see Sober and Wilson, *Unto Others* (Cambridge: Harvard, 1998) for their current view, which has changed somewhat.

II. OPENING THE BOXES

At this point, it should be clear why game-theoretic explanations of prosocial behaviors do not open up the black boxes to examine the proximate mechanisms inside. For as we have seen regarding the Stag Hunt game, all three styles of explanation—stability, assortation, and efficiency—predict the same observed behaviors in games. Furthermore, the Stag Hunt is not an example that is contrived to demonstrate this phenomenon. Any of the standard coordination games or bargaining games that are typically the subject of game-theoretic studies will behave the same way. Because the three standard explanantia in game-theoretic analyses will tend to agree across a wide range of models, we have the possibility of deriving explanations of great generality. That is, if all of the available proximate mechanisms have the same observational consequences, then there is little pressure to examine the details of those explanations, or uncover differences between them.

There is, however, another simplifying assumption in the game-theoretic models that has gone largely unnoticed.[13] For in each of the standard games—including the Stag Hunt, Prisoner's Dilemma, and Nash Bargaining Game[14]—each player faces the same available strategies and the same range of potential payoffs. In the language of game theory, we say that such games are symmetric. The restriction to symmetric games has an excellent pedigree. In his seminal paper, Nash made this restriction so as to express the assumption that the players have equal bargaining power. However, I am not aware of any recent discussions of the evolution of prosocial behaviors that explicitly justifies such an assumption. It is this restriction to symmetric games that has obscured several important conceptual and formal questions that need to be addressed.

II.1. Ambiguity in the Explanandum. Suppose that two individuals—call them *A* and *B*—must decide on how a resource is to be divided

---

[13] A notable exception may be found in Ken Binmore, *Game Theory and the Social Contract* (Cambridge: MIT, 1994).

[14] Also known as "Divide the Cake" or "Divide the Dollar."

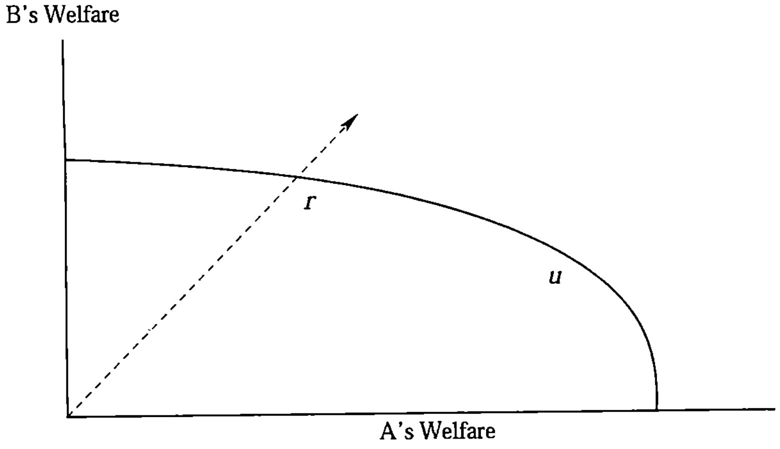


Figure 2: An asymmetric bargaining situation.

between them.[15] If we relax the assumption that their choices are symmetric, then we may diagram the situation as in Figure 2. There, the possible payoffs to *A* and to *B* are represented along the *X* and *Y* axes, respectively. Points further to the right represent outcomes giving *A* higher payoffs, while points higher along the *Y* axis represent outcomes giving *B* a high payoff. The range of jointly feasible outcomes is represented by the curve; any combination that would put the payoffs of *A* and *B* under the curve are feasible bargains. As is represented in Figure 2, player *A* has the potential to enjoy higher payoffs than player *B*. We need not specify why this is so—we can say that it is due to "differences in birth and natural endowment," as does John Rawls.[16]

The asymmetry of the game invites differences in opinion as to the "fair" or "just" bargain. The Rawlsian maximin solution is point *r*; for at point *r*, each player has an identical payoff, and no individual could do better without the other doing worse.[17] However, a utilitarian solution will identify point *u* as the just outcome, because that point maximizes the sum of the two payoffs.

Note that if the situation were symmetric, then point *r* and point *u* would be the same. Thus, the debate between the Rawlsian and the utilitarian does not arise in a game-theoretic analysis unless the game is asymmetric. Thus, if symmetric games are favored in discussions of the evolution of prosocial behaviors, we should expect a large degree of ambiguity as to exactly which prosocial behavior, norm, or convention we are examining.

---

[15] The form of this example is taken from Binmore.

[16] *A Theory of Justice* (Cambridge: Harvard, 1972).

[17] In general, if the range of feasible outcomes is a convex set, then the Rawslian solution may be found simply by drawing a line at a 45-degree angle through the origin, and intersecting it with the boundary of the set of feasible outcomes (where the boundary is also known as the Pareto frontier).

Indeed, it is not hard to find such ambiguity. Skyrms, for example, characterizes his project alternately as explaining the evolution of "justice," "fairness," and "the social contract." He also describes the explanandum as a sense of fairness and as a moral norm, although it is unclear that a moral "sense" can be equated with a norm.[18] The Prisoner's Dilemma—which may be the most influential game-theoretic model of prosocial behavior—has been understood alternately as a model of cooperation, fairness, and most recently as a model of altruism.[19] However, it should be obvious that fairness, cooperation, and altruism are not the same concepts. For instance, we may cooperate so that our individual self-interests are advanced, and this is not an altruistic behavior. Similarly, two uncooperative individuals may have a fair settlement imposed upon them. And although it is a case of cooperation when everyone agrees to drive on the right side of the road, there is nothing altruistic about that coordinated behavior.

The fact that these different concepts are frequently run together in game-theoretic analyses can be blamed largely upon the fact that the games in such discussions are typically symmetric. Already, if we use a formal tool like game theory to examine prosocial behaviors, we are necessarily subjecting the analysis to a good deal of simplification. But if we make the additional assumption that the players in such games are faced with identical choices and possible outcomes, then there is no longer any formal characteristic that can be used to pry apart alternate moral theories, or indeed, to distinguish between various moral concepts such as fairness, justice, cooperation, and altruism.

II.2. Formal Considerations. Not only is the explanandum obscured in models of symmetric games, but the explanans is also obscured to at least the same degree. To see this, we return to the example of an asymmetric bargaining situation illustrated in Figure 2. By showing how the asymmetric game allows us to distinguish among various explanantia, we may better see how symmetric games fail in this regard.

---

[18] Although it is important to note that Skyrms does not intend for his work to constitute a full-blown theory of the evolution of justice, or any other prosocial behavior or norm. As he says in *Evolution of the Social Contract*, his models are merely intended to be "the beginning of an explanation" of those concepts.

[19] See Axelrod; Binmore; and Sober and Wilson's *Unto Others*, respectively, for examples of these interpretations.

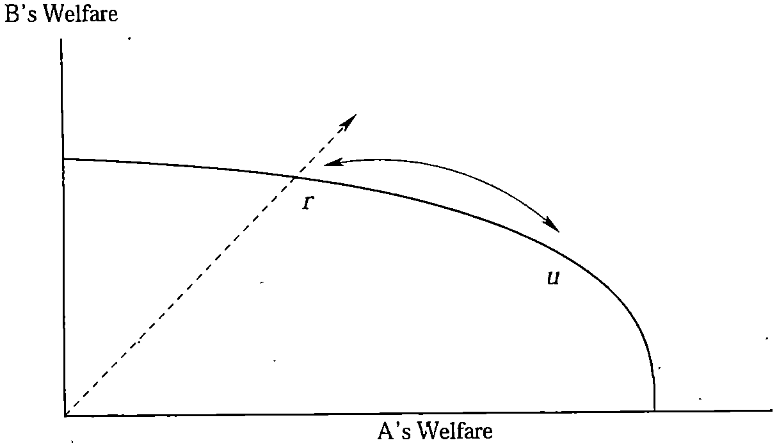


Figure 3: Group selection mechanisms should put the population somewhere between point *r* and point *u*, in the range indicated by the curved arrow.

For example, suppose that an interaction is asymmetric and occurs in the context of biological evolution and natural selection. As I have argued, if group selection mechanisms operate, then those mechanisms will tend to push the population to the efficient outcome. Because the efficient outcome in our simple example is the utilitarian solution at point *u*, group selection mechanisms will push the population toward that point. However, if the dominant mechanism is assortation, then the population dynamic will move the group's behavior along the dashed line toward the Rawlsian maximin solution at point *r*. If both mechanisms are operating to some significant degree, then it will turn out that neither point *u* nor point *r* will be stable. Instead, the stable point will be some compromise between those two points, somewhere along the curved line in Figure 3.

Even in this simple example, there are some surprising lessons that can be learned. The first concerns how we should formally model group selection mechanisms. Advocates of group selection—especially Sober and Wilson—place assortation mechanisms at the center of formal models of such processes. For group selection is said to operate to the degree to which different phenotypes become concentrated in different groups. For example, group selection will operate when there is some mechanism that tends to put altruists in one group, and selfish types in another. When such a process operates, the predominantly altruistic group will have a higher average fitness than the selfish group. Of course, if there is some form of intergroup selection operating, then the altruistic group will be favored.

In this way, assortation and group efficiency go hand-in-hand. Assortation brings about intergroup differences in group efficiency; group selection mechanisms select the more efficient group. It is uncontroversial that assortation mechanisms are correctly emphasized as a necessary prerequisite for group selection; and this is agreed upon even by those, like Richard Dawkins, who argue forcefully against group selectionist explanations.[20]

Although it is certainly correct that assortation and efficiency do work together in this way, it is an underappreciated fact that these two mechanisms may work against each other. For as I have argued above, in an asymmetric game, assortation mechanisms will push the population toward point *r*, while efficiency will push the population away from point *r*, and closer to point *u*. Thus, in an environment in which group selection operates, we should not expect to find the population at either point, but somewhere in a range between those two points, as in Figure 3.

Of course, if the game is symmetric, and point *u* is the same as point *r*, then it is safe to equate assortation and efficiency, putting them both under the general heading of "group selection." But since many interactions are not symmetric, there are many cases in which assortation and efficiency should be kept apart.

Precisely where the population will end up depends upon the relative influence of assortation and efficiency. For this reason, we must pay attention to the precise nature of the mechanisms that operate. This is in stark contrast to the conventional wisdom on this topic, which asserts that it does not matter precisely which assortation mechanism is operating. Both Skyrms and Sober and Wilson cite the same passage from Hamilton, that

> ...it obviously makes no difference if altruists settle with altruists because they are related...or because they recognize fellow altruists as such, or settle together because of some pleiotropic effect of the gene on habitat preference.[21]

In Hamilton's example, a gene causes altruistic behavior, and causes those with the gene to settle in a particular habitat. Thus, the altruistic individuals tend to settle in one habitat, while the selfish types settle in another.

In fact, assortation mechanisms are sometimes grouped together under the heading "greenbeard effect," so called because of a fanciful

---

[20] *The Selfish Gene* (New York: Oxford, 1976).

[21] Sober and Wilson, *Unto Others*, p. 134.

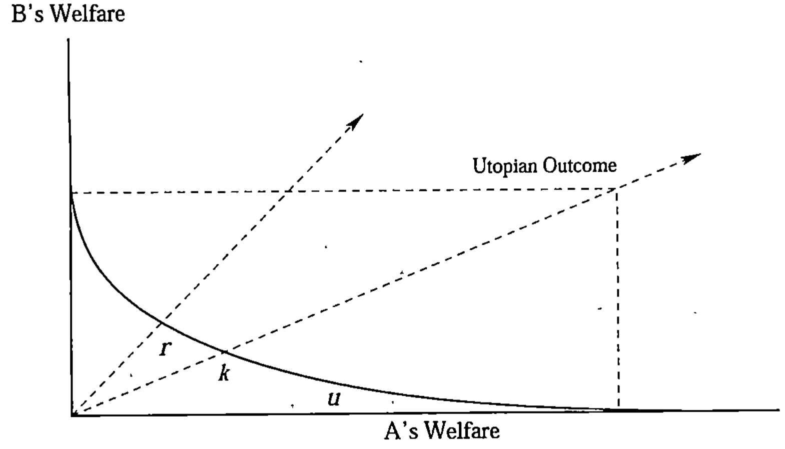


Figure 4: In a game without a convex payoff structure, the Kalai-Smorodinski, utilitarian, and Rawlsian outcomes may be distinct.

example due to Dawkins. In his example, altruists have green beards, and people with green beards prefer to interact with each other (*op. cit.*, p. 89). The point of this imaginary example of an assortation mechanism is that it does not matter what mechanism sorts the population into altruist and nonaltruist groups, so long as something does.

The considerations that I have raised above suggest, however, that we should be hesitant to treat all mechanisms of assortment the same. If different mechanisms of assortment operate to different degrees, then the population will be pushed further toward point *r* (if assortment is strong relative to efficiency), or closer to point *u*.[22]

II.3. Ambiguity among Proximate Mechanisms. The points *u* and *r* correspond to two different moral intuitions—the utilitarian's intuition that the fair outcome maximizes the welfare of everyone, and the Rawlsian intuition that we ought to maximize the welfare of the least well-off. Other solution concepts may be illustrated by other asymmetric payoff curves. The payoff curve in Figure 2 is said to be "convex," because a line connecting any points inside the curve does not go outside the curve. If this restriction is lifted, then we have the possibility of considering games whose payoff structure like that depicted in Figure 4. Clearly, the utilitarian and Rawlsian solutions may be arbitrarily far apart in a game with a concave payoff structure. Additionally, it is easy to illustrate other conceptions of "fair" outcomes, such as the Kalai-Smorodinksi solution. According to Kalai-Smorodinski, the fair outcome is calculated by first locating the Utopian solution, which is a state in which *A* and *B* both enjoy their maximum possible payoffs. Typically, that point will lie outside the range of possible outcomes, because *A* and *B* cannot simultaneously obtain their highest possible payoffs. Accordingly, the Kalai-Smorodinksi solution concept tells us to draw a straight line from the status quo point (or original position) to the Utopian outcome.[23] The fair outcome is the best point on the line inside the range of feasible outcomes. This is illustrated by point *k* in Figure 4.

---

[22] The conflict between Rawlsian and utilitarian solution concepts in a biological setting corresponds closely to the conflict between the evolution of female-biased sex ratios and the evolution of sex ratios approaching unity. For in a group-selection model, within-group selection will favor the evolution of a 1:1 sex ratio (as in R. A. Fisher's original argument), while between-group selection will favor a female-biased sex ratio. In a group-selection model, the two pressures are in conflict with each other; accordingly, we observe that the actual sex ratios are a compromise between those two pressures. Specifically, the stronger the between-group selection, the more heavily biased toward a female sex ratio. A female-biased sex ratio produces offspring more efficiently; thus, the tendency toward a female-biased sex ratio corresponds to the efficient utilitarian solution. On the other hand, a 1:1 sex ratio corresponds to (what I have called) the Rawlsian outcome. Accordingly, the mathematics of a group-selectionist model of sex ratio evolution is analogous to the mathematics of asymmetric games under the dynamics I have informally described above. For a discussion of the evolution of female-biased sex ratios, see D.S. Wilson and R.K. Colwell, "Evolution of Sex Ratio in Structured Demes," *Evolution*, xxxv (1981): 882–97.

Clearly, each solution concept enjoys a significant degree of intuitive plausibility. Because different mechanisms—both biological and social—have the potential to push populations toward different solution concepts, it is important to clarify the explanandum when we set out to explain "the origins of prosocial behavior." Instead of trying to understand the origins of a broad range of prosocial practices, intuitions, and sentiments, we should be setting out to explain the origins of particular moral intuitions or practices. For instance, we may try to explain the origins of broadly utilitarian sentiments, or we might try to explain why the maximin principle appeals to our sense of fair play.

In fact, asymmetric games allow us to highlight the different causal influences of various social and biological mechanisms. If a significant proportion of morally relevant interactions are asymmetric, and these causal factors trend toward different solution concepts, then we should expect that "our moral intuitions" are not a unified whole. Instead, our moral sentiments and practices would be the result of overlapping and contradictory social and biological influences.

---

[23] The Kalai-Smorodinski solution plays an important role in the theory advocated by Binmore.

Indeed, when experimental economists conduct studies of human beings' actual social behavior in bargaining situations, they model our behavior as governed by an overlapping set of conflicting norms. It should come as no shock to either conventional wisdom or moral theory that our moral sentiments, intuitions, and social practices are highly conflicted. Various environmental triggers activate different norms to varying degrees. These triggers have been investigated in detail by experimental economists, who have identified the roles played by institutions, the scale of the stakes involved, as well as cultural differences.[24] These norms may conflict with each other, and each is activated to varying degrees by different environmental cues. For example, Elizabeth Hoffman and others have argued that humans calculate the so-called "social distance" between themselves and their partners.[25] The smaller the social distance, the greater the degree of activation of norms favoring equal division in bargaining situations. Accordingly, when one's bargaining partner is a member of one's own social group, test subjects overwhelmingly tend toward equal division, even when there is a financial disincentive to one partner.

III. LESSONS LEARNED

If our moral intuitions are conflicted, and are triggered by a variety of different factors, then it is a mistake to take on the project of explaining "the origins of fairness" —or any other general prosocial behavior—simpliciter. For if the project is specified in such a way, then it incorrectly treats our moral intuitions as if they were a unified and consistent set. This mistake is compounded if it also commits us to treating all evolutionary (both social and biological) mechanisms as if they can all be described by the same dynamics.

It is an interesting change that the sociobiologists, including E.O. Wilson[26] and Michael Ruse,[27] did not think of their project as explaining the origins of fairness, altruism, justice, the social contract, or any other general moral category. Instead, their proposed explanations were restricted to particular social practices such as inheritance, racism, preferential treatment of genetic relatives, and so on. In contrast, current work on the evolution of prosocial behaviors tends to be much more general. Skyrms, J. McKenzie Alexander,[28] Sober, Wilson, and others tend to be concerned with general moral categories like "the social contract," "altruism," "justice," and "fairness."

---

[24] See Bruno S. Frey and Iris Bohnet, "Institutions Affect Fairness: Experimental Investigations," *Behavior,* LXXV (1980): 262–300; Lisa A. Cameron, "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry,* XXXVII (1999): 47–59; Alvin Roth, V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review,* LXXXI (1991): 1068–95.

[25] Hoffman, Kevin McCabe, Leoth Shachat, and Vernon Smith, "Preferences, Property Rights and Anonymity in Bargaining Games," *Games and Economic Behavior,* VII (1994): 346–80; Hoffman, McCabe, and Smith, "Social Distance and Other-Regarding Behavior in Dictator Games," *The American Economic Review,* LXXXVI (1996): 653–60.

[26] *Sociobiology: The New Synthesis* (Cambridge: Harvard, 1975).

[27] Ruse and Wilson, "Moral Philosophy as Applied Science," in Sober, ed., *Conceptual Issues in Evolutionary Biology* (Cambridge: MIT, 1984), pp. 421–38.

The switch from specific and particularized explanations to highly general explanations coincides with the rise of game theory as the favored explanatory tool. If the arguments above are correct, then it is not a coincidence that the explanandum changed in this way when game theory became widely employed. After all, the standard games used to model prosocial behaviors are symmetric. Because all of these moral concepts and prosocial behaviors are conflated in symmetric games, the models encourage us to conflate all of these moral concepts. Similarly, the mechanisms that may lead to prosocial behaviors also push populations to the same outcomes, provided that the games are symmetric. Thus, the exclusive reliance on symmetric games encourages us to ignore the underlying mechanisms that are ultimately responsible for the evolution of prosocial behaviors.

There is a cost of ignoring the specific mechanisms that are supposed to lead to prosocial behaviors. Any such proposed explanation is vulnerable to the criticism that the devil is in the details. That is, if one can come up with a situation in which there is some force that is significantly different from the standard game-theoretic models, and which leads to a different outcome, then one has a substantive criticism. And of course, this is the approach that is standardly taken to object to game-theoretic explanations of prosocial behaviors. For instance, D'Arms and others have pointed out that there are cases in which we might expect an anti-correlation in interaction[29]; Philip Kitcher has argued that the presence or absence of coalitions may play an important role in biological contexts[30]; Neil Tennant has argued that it may make a difference whether reproduction is sexual or asexual in biological models[31]; and I have questioned elsewhere the assumption that biological and social evolution are analogous.[32]

---

[28] "Evolutionary Explanations of Distributive Justice," *Philosophy of Science*, LXVII (2000): 490–516.

[29] "Game Theoretic Explanations and the Evolution of Justice."

[30] "Games Social Animals Play: Commentary on Brian Skyrms's *Evolution of the Social Contract*," *Philosophy and Phenomenological Research*, LIX (1999): 221–28.

[31] "Sex and the Evolution of Fair-dealing," *Philosophy of Science*, LXVI (1999): 391–414.

[32] Zachary Ernst, "Explaining the Social Contract," *British Journal for the Philosophy of Science*, LII (2001): 1–24.

This line of objection is basically sound. It is too much to expect a single model to be agnostic, not only about which mechanism is supposed to shoulder the explanatory burden, but also about what the explanandum is. Accordingly, it is relatively easy to identify potential causal influences that may adversely affect the explanatory significance of the game-theoretic models. But we should not take this line of objection to be more significant than it actually is. For there is no intrinsic feature of the game-theoretic models that forces us to neglect specifying the explanandum. In order to deal with these criticisms, game-theoretic models must be detailed enough to reflect the finer distinctions between various evolutionary mechanisms as well as the differences between varieties of prosocial behaviors. Accordingly, we must pay attention to asymmetric games.

ZACHARY ERNST

Florida State University

ON THE POSSIBILITY OF PARETIAN EGALITARIANISM*

We here address the question of how, for a theory of justice, a concern for the promotion of equality can be combined with a concern for making people as well off as possible. Leximin, which requires making the worst off position as well off as possible, is one way of combining a concern for making people's lives go well with a special concern for those who are especially poorly off. Many egalitarians, however, reject its near monomaniacal focus on the worst off position (to the exclusion of other poorly off persons). In this paper, we explore the possibility of combining a weak kind of egalitarianism with a weak kind of efficiency requirement in a way that avoids leximin's obsession with the worst off position. For example, one may consider solving all cases where efficiency is not at issue by choosing the alternative that is most equal according to the Gini-coefficient or some other well-established inequality measure. All standard inequality measures sometimes judge an alternative as more equal than another alternative even though the latter maximizes the benefits of the worst off. Thus it may seem like a promising way of avoiding the leximin approach within an egalitarian framework. Surprisingly, given certain generally accepted assumptions, this turns out to be impossible. The only possible way of combining weak egalitarianism with weak efficiency requires, we shall show, the rejection of a widely accepted—but perhaps dubious—contraction consistency condition on justice or the acceptance of some version of the leximin principle.

I. THE GENERAL FRAMEWORK

To specify fully an egalitarian theory one must specify the type of benefits that it seeks to equalize. Throughout, we leave open the relevant conception of benefit (resources, primary goods, brute luck, well-being, and the like). References to a person being worse off than another should be understood in terms of the relevant benefits.

We shall assume, for the sake of argument, that benefits are fully measurable and interpersonally comparable. This may seem like a strong assumption, but in the present context it is a very weak assumption. The assumption that benefits are so measurable and comparable

---

* For helpful comments, we thank Nils Holtug, Luc Lauwers, Kasper Lippert-Rasmussen, Ashley Piggins, Alex Voorhoeve, and Paul Weirich.

0022-362X/05/0203/126–54

does not entail that such information is *relevant* for the moral assessment of options. The assumption is simply that such information is available. This ensures that no principle of justice is ruled out merely on the grounds that it presupposes that benefits are measurable or comparable in ways that they are not. One important qualification, however, is needed here: if benefits are not—as required by leximin—ordinally interpersonally comparable (that is, if there are no facts about who is worse off than whom), then leximin is not possible. Because we are exploring the possibility of avoiding leximin (or something like it) where it is possible, the assumption that benefits are at least so measurable and comparable is essential to our results.

We shall be concerned with the assessment of the justice of alternatives, where alternatives are possible objects of choice (for example, actions or social policies). Alternatives may have all kinds of features: they generate a certain distribution of benefits, satisfy or violate various rights, involve various intentions, and so on. In what follows, we shall assume that the only relevant information for the assessment of justice is the benefit distribution that an alternative generates. More formally, we shall assume:

> *Benefitism*: Alternatives can be identified with (and thus their justice assessed solely on the basis of) their benefit distributions.

Benefitism is a generalization of welfarism. Although it does not assume that welfare is all that matters, it does assume that justice supervenes on individual benefits. If two alternatives generate the same distribution of benefits, then they have the same status with respect to justice.

Benefitism is a very strong assumption. It rules out the relevance, for example, of respecting the wills of agents (for example, as reflected in their consent). It holds that there is no difference with respect to justice between forcing a person to go jogging against her will and merely successfully encouraging her to do so, if the benefits to her and everyone else are the same. Although we would reject this condition for this reason, we shall here grant it, since it is not relevant for the issue at hand. The issue of whether something like leximin is the only plausible egalitarian theory of justice arises even if equality promotion is limited, for example, to those alternatives that respect the wills of agents (for example, violate no rights). Granting Benefitism will simplify the presentation. The results could, however, be recast as applying where the only relevant differences between alternatives concern the distributions of benefit that they generate.

Given Benefitism, we can identify an alternative with the benefit distribution that it generates, and in what follows we shall do so for simplicity. We shall further assume that the set of benefit distributions generated by the set of possible alternatives is *rich* in the following sense:

> *Domain Richness*: For any logically possible benefit distribution there is an alternative that generates that distribution.

This condition rules out, for example, the possibility that, where there are just three people, the distribution <3,7,9> (3 to the first person, 7 to the second, 9 to the third) is not one of the alternatives. All logically possible benefit distributions are among the alternatives. This is not to say that all are part of any given *feasible* set (the alternatives that are open to an agent on a given occasion). Of course, there are lots of logically possible benefit distributions that are not feasible on a given occasion. The claim here is about the range of benefit distributions that can be assessed by justice, equality, or other normatively relevant perspectives. The condition holds that such judgments can be made for all logically possible distributions.

We believe that this is a highly plausible condition. Benefit distributions here play the role of test cases for a theory of justice. All logically possible test cases—assuming, as we shall, a finite population—are admissible.

Justice can be understood in *axiological* terms—what is at least as just as what—or in *deontic* terms—what is just (permitted by justice) relative to a set of feasible alternatives. Deontic justice in this sense need not be grounded in deontological properties. It might be, for example, that an option is deontically just if and only if it maximizes total well-being. We shall start by considering axiological justice and then turn to deontic justice.

Axiological justice is concerned with the *justice relation* of (one alternative) being-at-least-as-just-as (another). Following the standard definitions, (1) an alternative is *more just* than another if and only if it is at least as just and the other is not at least as just as it; and (2) an alternative is *equally as just* as another if and only if it is at least as just and that other is also at least as just as it.

II. PARETIAN EGALITARIANISM

Our general question is: What kinds of egalitarianism are compatible with the view that increasing everyone's benefits makes things more just? *Pure egalitarianism* holds that justice is concerned only with the equality of the distribution of benefits. Justice is sometimes understood as the *purely comparative* concern for ensuring that people get what they are owed, and, so understood, pure egalitarianism applied to brute luck benefits (benefits that were not deliberately produced by anyone) is not an implausible view. From a purely comparative perspective, it is worse when some get more brute luck benefits than others. Here, however, we are concerned with justice understood as the concern for people getting what they are owed in a *non-purely comparative manner*, and part of what individuals are so owed, we shall assume, is that their benefits be as high as possible, given the benefits that others receive. As a theory of justice so understood, pure egalitarianism is implausible, because it holds, for example, that perfect equality (for example, <5,5,5>) is more just than everyone having more benefits but with an unequal distribution (for example, <6,9,12>). Pure egalitarianism fails, that is, to recognize that justice is concerned in part with increasing the benefits that individuals obtain.

This problem with pure egalitarianism has lead many egalitarians to embrace *leximin*, which holds that justice is increased if and only if (1) the benefits to the worst off position are increased, or, (2) in case of ties, the benefits to the second worse off position are increased, and so on. Leximin adequately recognizes that justice is concerned with increasing people's benefits, but it suffers from a different problem. It gives *absolute* priority to the worst off position. It holds that justice can require giving a very small benefit to the worst off position rather than giving large benefits to many slightly better off people who would still have less than average benefits (for example, that <2,2,2,2,400> is more just than <1,80,80,80,400>).

Leximin is but one of a class of principles that give strict priority to benefits to the worst off position. In what follows, we shall be concerned with *weak maximin*, which is the principle that an alternative is more just if it makes the worst off position better off. Weak maximin is silent about cases where the worst off position is equally well off (for example, <2,3,5> versus <2,4,4>), and is compatible with various ways of handling such cases. In what follows we shall examine the possibility of avoiding the strict priority of benefits to the worst off required by weak maximin.

A different—and seemingly more promising—way of combining a concern for the worse off with a concern for increasing people's benefits is (axiological) *Paretian Egalitarianism*, which is the conjunction of Strong Pareto and Weak Egalitarianism, as defined below.

> *Strong Pareto*: For any two alternatives x and y, if each person has at least as much benefits in x as in y, then (1) x is as least as just as y, and (2) if there is at least one person that has more benefits in x than in y, then x is more just than y.

This is a weak efficiency condition on the promotion of benefits (much weaker than the utilitarian sum-total conception of efficiency).

It requires, for example, that <2,4,6> be judged more just than <1,4,6> and also more just than <2,3,6>. It is silent about whether <2,4,6> is more just than <99,1,6>.

For the next condition, we appeal to the equality of distributions (which we will discuss in detail in the next section). We shall appeal to a relation that ranks distributions in terms of how equal they are (for example, that ranks <2,2,2> as more equal than <1,2,3>). Moreover, we need to introduce the following definition. Call a distribution *anonymously Pareto incomparable* to another just in case it, and each of its benefit permutations (that is, the same distribution of benefits except perhaps with people occupying different positions in the distribution), is Pareto incomparable (that is, better for some and worse for some others) to the other distribution. This is a strong kind of Pareto incomparability. For example, <1,3,6> is Pareto incomparable with <2,1,6> but it is not anonymously Pareto incomparable, since <1,2,6> is a permutation of the latter and it is Pareto inferior to <1,3,6>.

Consider now the following condition:

> *Weak Egalitarianism*: For any two alternatives that are anonymously Pareto incomparable, (1) if their two benefit distributions are equally equal, then they are equally just, and (2) if the benefit distribution of one is more equal than that of the other, then the former is more just.

This is a very weak egalitarian condition. It does not require (as pure egalitarianism does) that equality be the only concern of justice. It only holds that equality is determinative of justice in the case where two alternatives are anonymously Pareto incomparable. It is a very weak egalitarian principle, because it is silent even in many cases of Pareto incomparability (namely, those that are not anonymously Pareto incomparable).

Weak Egalitarianism could be rejected for several possible reasons. One is the denial that equality is ever relevant to justice. A second reason allows that equality is relevant but only as a pro tanto consideration that can be outweighed by other considerations (for example, sufficient increases in total benefits). A third reason holds that equality is indeed determinative of justice under certain conditions, but these conditions are limited to cases where the total is the same. Equality, on this view, is simply a tie-breaker when the totals are the same.

We believe that equality is more central to justice than these objections allow. In any case, we want to explore here whether weak maximin is forced on a certain kind of weak egalitarianism, and hence we shall accept this condition. The conjunction of Strong Pareto and Weak Egalitarianism defines *Paretian Egalitarianism.* There are different forms of Paretian Egalitarianism, because the conjunction of these two conditions leaves open how distributions are ranked that are Pareto incomparable but not anonymously Pareto incomparable (for example, <8,10,12> versus <9,8,12>). Different forms of Paretian Egalitarianism fill in these cases in different ways.

One version of Paretian Egalitarianism that we shall consider below is *Paretian Moderate Egalitarianism*, which is defined as the conjunction of Strong Pareto and the following stronger equality condition:

> *Moderate Egalitarianism*: For any two alternatives that are Pareto incomparable, (1) if their two benefit distributions are equally equal, then they are equally just, and (2) if the benefit distribution of one is more equal than that of the other, then the former is more just.

This is just like Weak Egalitarianism except that it makes equality determinative of justice whenever two alternatives are Pareto incomparable, and not merely when they are anonymously Pareto incomparable. For example, if <9,8,8> is more equal than <8,10,12>, then it says that it is also more just, whereas Weak Egalitarianism is silent (since they are not anonymously Pareto incomparable). Moderate Egalitarianism gives equality the strongest role possible while respecting Strong Pareto. It entails Weak Egalitarianism, but not vice versa.

Paretian Egalitarianism seems like a promising way of combining equality with the efficient promotion of benefits while avoiding weak maximin's obsession with the worst off individual. Paretian Egalitarians, for example, may wish to hold that <2,10,100> is less equal than <1,100,100>—and hence less just (given that they are anonymously Pareto incomparable). We shall now see, however, that if certain seemingly intuitively plausible assumptions are satisfied, then Paretian Moderate Egalitarianism is impossible, and Paretian Egalitarianism cannot disagree with weak maximin. It cannot, for example, hold that <2,10,100> is less just—or even less equal—than <1,100,100>. Thus, Paretian Egalitarianism appears not to be a real alternative to weak maximin.[1]

---

[1] We are here building upon Bertil Tungodden, "Egalitarianism: Is Leximin the Only Option," *Economics and Philosophy*, XVI (2000): 229–45, "The Value of Equality," *Economics and Philosophy*, XIX (2003): 1–44, and "Hammond Equity: A Generalization," Discussion Paper, Norwegian School of Economics and Business Administration (2001). Our conditions are cast in terms of Paretian egalitarianism and the equality relation rather than in terms of a general justice relation (as Tungodden does). Tungodden's work builds upon that of Peter Hammond, "Equity, Arrow's Condition, and Rawls' Difference Principle," *Econometrica*, XLIV (1976): 793–804, and "Equality in Two-Person Situations—Some Consequences," *Econometrica*, XLIV (1979): 1127–35.

III. THE EQUALITY RELATION

Paretian egalitarianism appeals to the *equality relation* of one distribution of benefits being at least as equal as that of another. Following the standard definitions, (1) an alternative is *more equal* than another if and only if it is at least as equal and the other is not at least as equal as it; and (2) an alternative is *equally as equal* as another if and only if it is at least as equal and that other is also at least as equal as it. We here identify some uncontroversial assumptions that we make about this relation.

Our framework assumption of Domain Richness is an assumption about the range of benefit distributions that can be assessed by justice, equality, or other normatively relevant perspectives, and thus applies for the equality relation. It thus entails that the domain of the equality relation is the set of all logically possible distributions of benefits. We do not assume that the equality relation is complete, nor, except for one result at the end, that it satisfies any consistency requirement (such as acyclicity).

We shall, however, make the following three assumptions about this relation, each of which is at least close to being a conceptual truth about equality.

Consider, first:

> *Perfect Equality*: A distribution in which everyone gets the same benefits is more equal than one in which this is not so.

This is completely uncontroversial. It says, for example, that <3,3,3> is more equal than <2,3,4> and more equal than <5,6,6>.

A second condition is:

> *Equality Strong Anonymity*: For any two alternatives, x and y, if x is at least as equal as y, then, for any alternative, x^*, for which the benefit distribution is a permutation of the benefit distribution of x (that is, has the same pattern of benefits, but perhaps with individuals receiving different allocations), x^* is at least as equal as y.

This is a limited anonymity requirement on equality. It requires that, if <2,5,13> is more equal than <2,9,12>, then so is <5,2,13>. Equality Strong Anonymity is highly plausible, since equality is based solely on the pattern of benefits and is not concerned with who gets what.

A third plausible condition is:

> *Strong Contracting Extremes Equality*: If (1a) the benefits of all the best off individuals are reduced without making them cease to the best off, and/ or (1b) the benefits of all the worst off individuals are increased without

making them cease to be the worst off, and (2) no one else is affected, then the result is more equal than the original distribution.[2]

This condition is also extremely plausible. It holds, for example, that <2,2,5,6,8,8> is more equal than each of <1,1,5,6,9,9>, <2,2,5,6,9,9>, and <1,1,5,6,8,8>.

We are almost ready to begin examining the possibility of Paretian Egalitarianism disagreeing with weak maximin. First, however, it will be useful to note that, *where there are only two people,* weak maximin and the equality relation are intimately related.

> *Result 1*: If (1) we have Benefitism and Domain Richness, (2) the equality relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, and (3) there are only two people, then for any two anonymously Pareto incomparable alternatives, the alternative that makes the worst off position better off is more equal.

The proofs of this and the following results are located in the appendix. We shall simply illustrate in our main text the core nature of each result. Consider here the anonymously Pareto incomparable pair <4,2> and <1,5>. The result establishes the intuitively obvious judgment that the former is more equal. Strong Contracting Extremes Equality entails that <2,4> is more equal than <1,5>, and hence Equality Strong Anonymity entails that <4,2> also is.

In the two person case, then, equality always favors the worst off when alternatives are anonymously Pareto incomparable. This ensures that, in the two person case, Weak Egalitarianism never conflicts with weak maximin. If two alternatives are not anonymously Pareto incomparable, then Weak Egalitarianism is silent and no conflict is possible. If the two alternatives are anonymously Pareto comparable, then equality agrees with weak maximin. This need not be the case, however, when there are more than two persons. The conjunction of Strong Contracting Extremes and Equality Strong Anonymity are silent, for example, about whether <6,8,100> is more equal than <100,5,100>. Nonetheless, as we shall now show, if certain additional assumptions are made, then Paretian Egalitarianism must hold that <6,8,100> is more just—and indeed, more equal—than <100,5,100>. More generally, even in the multi-person case, Paretian Egalitarianism, like weak maximin, must not judge an alternative as at least as just as another if it is worse for the worst off position.

---

[2] This condition is stronger than that used by Tungodden in his cited papers. His version of the condition requires that antecedent clauses (1a) and (1b) both hold. The condition was first formulated by Peter Vallentyne, "Equality, Efficiency, and Priority for the Worse Off," *Economics and Philosophy*, XVI (2000): 1–19. It there takes (roughly) the stronger form that we invoke here. For brevity, we leave implicit the following antecedent clause: there is no natural zero for benefits (no nonarbitrary way of drawing the line between positive and negative levels), or all benefit levels are positive, or all are negative. This clause is needed so as not to require, for example, that <0,10> is less equal than <0,1>. Some measures of equality hold that these two distributions are equally equal on the grounds that each person has the same proportion of the total benefits under the first as under the second. Although we believe that Strong Contracting Extremes Equality is plausible without being conditional on this antecedent clause, we do not presuppose this in what follows. Our proofs and examples are compatible with the weaker conditional version of the condition.

IV. AXIOLOGICAL JUSTICE

Axiological justice is concerned with the justice relation, and hence we need to specify the structure of this relation. We shall assume that the justice relation satisfies the following consistency condition:

> *Acyclicity*: If, for alternatives $x_1,\ldots, x_n$, x_1 is more just than x_2, x_2 is more just than $x_3,\ldots$, and x_{n-1} is more just than x_n, then x_n is not more just than x_1.

Acyclicity is much weaker than the better known requirement of transitivity. It applies only to chains of where each alternative is more just than its successor (as opposed to being at least as just) and allows the possibilities of silence (no ranking) and of holding the first alternative be judged equally just as the last (whereas transitivity requires that it be judged more just). It is about as close to being uncontroversial as one can get when it comes to consistency requirements on the justice relation.

For some of the results, we will strengthen this condition slightly in two ways. One is to invoke:

> *Consistency*: If, for alternatives $x_1,\ldots, x_n$, x_1 is at least as just as x_2, x_2 is at least as just as $x_3,\ldots$, and x_{n-1} is at least as just as x_n, then (1) x_n is not more just than x_1, and (2) if for some i inclusively between 1 and n–1, x_i is more just than x_{i+1}, then x_n is not at least as just as x_1.[3]

This strengthens Acyclicity by covering chains where each alternative is at least as just (as opposed to more just) as its successor. If the pairs are equally just, then it concludes that the last alternative is *not more just* than the first (as opposed to the silence of Acyclicity). Moreover, if all the relations in this chain are the relations of being more just, then Consistency strengthens Acyclicity by requiring that the last alternative *not be at least as just* as the first (as opposed to Acyclicity's requirement that it not be more just). Like Acyclicity, Consistency is accepted by almost everyone.

---

[3] This condition comes from Kotaro Suzumura, "Remarks on the Theory of Collective Choice," *Economica*, XLIII (1976): 381–90. He proves that Consistency is a necessary and sufficient condition for a reflexive ranking relation to be extendible to a complete, transitive, and reflexive relation.

We will also consider the implications of the following strengthening of Acyclicity:

> *Quasi-Transitivity*: If, for alternatives x_1, x_2, and x_3, x_1 is more just than x_2, x_2 is more just than x_3, then x_1 is more just than x_3.

This is strictly stronger than Acyclicity in that it requires that the first element be more just than the last (as opposed to Acyclicity's requirement that the last is not more just than the first). Quasi-Transitivity is strictly weaker than transitivity, since (like Acyclicity) it says nothing about chains in which one or more alternatives is only ranked as at least as just as its successor (as opposed to more just). It is also weaker than Consistency in this respect. It is, however, stronger than Consistency in another respect: for chains of alternatives that are each more just than their successors, it requires that the first be more just than the last, whereas Consistency allows silence.

Quasi-Transitivity is much more controversial than Consistency. Indeed, one of us would reject it on the ground that sometimes silence is appropriate concerning the ranking of the first and last alternatives. Fortunately, our core results do not depend on this assumption. We shall merely note how they can be strengthened if one assumes Quasi-Transitivity as opposed to Consistency.

It is important to note that we do not assume that the justice relation is complete (that is, that for any two alternatives, at least one is at least as just as the other). Paretian Egalitarians, for example, may remain silent about the ranking of two alternatives which are Pareto incomparable (and hence not covered by Strong Pareto) but not anonymously so (and hence not covered by Weak Egalitarianism). Moreover, even in cases of anonymous Pareto incomparability, Weak Egalitarianism may be silent because the equality relation may be incomplete and be silent about which alternative is more equal. This possible incompleteness of the justice relation is unproblematic for our analysis. Our results only rely on statements about the justice relation and the equality relation in uncontroversial cases.

We are now ready to present our first result concerning Paretian Egalitarianism. Interestingly, given the weak and plausible conditions invoked, it turns out that Paretian Moderate Egalitarianism is impossible.[4]

---

[4] It is worth noting the connection of this result with the impossibility of combining Moderate Egalitarianism with the Suppes-Sen condition—according to which x is more just than y if it is anonymously Pareto superior to y. This latter impossibility is

Result 2: If (1) we have Benefitism and Domain Richness, (2) the justice relation satisfies Acyclicity, and (3) the equality relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, then Paretian Moderate Egalitarianism is impossible.

The force of this result (and the structure of the proof) can be seen by considering <7,8,11>, <10,8,7>, and <7,8,9>. Note that Strong Contracting Extremes Equality and Equality Strong Anonymity jointly entail that <7,8,9> is more equal than <10,8,7>, which in turn is more equal than <7,8,11>. Paretian Moderate Egalitarianism makes the following judgments: (1) <7,8,11> is more just than <7,8,9> (by Strong Pareto), (2) <7,8,9> is more just than <10,8,7> (because the two are Pareto incomparable, <7,8,9> is more equal), and (3) <10,8,7> is more just than <7,8,11> (because the two are Pareto incomparable, <10,8,7> is more equal). These three judgments, however, violate Acyclicity, which, based on the first two judgments, requires that <10,8,7> *not* be judged more just than <7,8,11>. Paretian Moderate Egalitarianism violates Acyclicity, and thus is impossible if Acyclicity is required, as we believe it should be.

Assuming Acyclicity and the other conditions, Paretian Moderate Egalitarianism is impossible. Paretian Moderate Egalitarianism is not, however, the only form of Paretian Egalitarianism. The following result, however, establishes that no form of Paretian Egalitarianism can significantly disagree with weak maximin.

Result 3: If (1) we have Benefitism and Domain Richness, (2) the justice relation satisfies Acyclicity, and (3) the equality relation satisfies Perfect Equality, then Paretian Egalitarianism judges that a distribution is *not less just* than another if it gives greater benefits to the worst off position. Moreover, (a) if the justice relation also satisfies Consistency, then Paretian Egalitarianism judges that a distribution is *not equally or less just*, if it gives greater benefits to the worst off position, and (b) if the justice relation also satisfies Quasi-Transitivity, then Paretian Egalitarianism judges that a distribution is *more just* than another if it gives greater benefits to the worst off position.

---

straightforward, given that Moderate Egalitarianism requires that x be less just than y, whenever (1) x is anonymously Pareto superior to y but not Pareto superior, and (2) x is less equal than y (for example, x = <2,10> and y = <9,2>). Although our conditions do not entail Suppes-Sen (because we do not assume that the justice relation is anonymous), we do assume something close to it. Equality Strong Anonymity and Paretian Moderate Egalitarianism entail (assuming the reflexivity of equality) Justice Anonymity (a distribution and any permutation are equally just), and Justice Anonymity and Strong Pareto entail Suppes-Sen, if one assumes transitivity. We assume, however, only Acyclicity (which is not sufficient for the entailment).

To see the force of this result, consider <3,3,99>, <2,4,4>, and <3,3,3>. Paretian Egalitarianism makes the following judgments: (1) <3,3,99> is more just than <3,3,3> (by Strong Pareto), and (2) <3,3,3> is more just than <2,4,4> (because they are anonymously Pareto incomparable and Perfect Equality entails that <3,3,3> is more equal). Acyclicity then entails that <3,3,99> is not less just than <2,4,4>. Moreover, Consistency entails that it is not equally or less just, and Quasi-Transitivity entails that it is more just. Thus, given the above conditions, Paretian Egalitarianism cannot judge <2,4,4> to be more equal than <3,3,99> (since if it did, given that they are anonymously Pareto incomparable, <2,4,4> would be judged more just, and that contradicts the results just obtained). Equality, that is, cannot be measured in a way that significantly diverges from weak maximin.

Given that Acyclicity is uncontroversial, the result establishes that Paretian Egalitarianism cannot judge an alternative to be *more just* than another if it is worse for the worst off position. This allows, however, that it might judge such an alternative to be *equally as just* as the other. Given that Consistency is also uncontroversial, the results also rule out this possibility. Thus, Paretian Egalitarianism cannot contradict weak maximin. It may be silent in certain cases where one alternative is better for the worst off than another, but it cannot affirm that the second is equally, or more, just than the first. This is enough, it seems, to eliminate any hope of Paretian Egalitarianism being significantly distinct from weak maximin. The much more controversial Quasi-Transitivity is not needed to establish this conclusion—although, if granted, it significantly strengthens the conclusion by ruling out the possibility of silence and thus requiring that Paretian Egalitarianism judge more just an alternative that is better for the worst off position.

It is worth noting that, assuming Consistency, the above result establishes that the measurement of *equality*—and not merely justice—must not contradict weak maximin *where the alternatives are anonymously Pareto incomparable*. This is because Paretian Egalitarianism ranks the justice of such distributions on the basis of their equality. The result does not require that equality agree with weak maximin in other cases. For example, it does not require that <2,9> is more equal than <8,1>, since the two options are not anonymously Pareto incomparable.

It is also worth noting that the conditions of the above result are compatible. It is possible, that is, to satisfy all of them. Leximin, for example, satisfies all of them. Indeed, this is true of all of the results of this paper, and thus we shall not bother to repeat the reminder

University Library ALIPORE CAMPUS

that the conditions invoked are jointly compatible (and satisfied by leximin).

There is, however, a possibility that we have not yet explored. We have so far focused on the justice relation. We have, that is, focused on *axiological* justice (what is more just than what). A different approach is to formulate the conditions of justice in deontic terms, that is, in terms of what is just (that is, permitted by justice) relative to set of feasible alternatives. This latter approach does not attempt to provide a global ranking of alternatives. Instead, it attempts simply to determine which of any given set of feasible alternatives are just. We shall now explore whether the above results remain valid in this context.[5]

V. DEONTIC JUSTICE

To start, let us reformulate the above axiological conditions that were used to define Paretian Egalitarianism and the two specific versions. The following definitions will be used below. An alternative is *Pareto optimal*, relative to a given feasible set, if and only if no feasible alternative is Pareto superior (that is, makes some better off and no one worse off). An alternative is *anonymously Pareto optimal*, relative to a given feasible set, if and only if it is *feasible* and neither it, *nor any of its permutations*, is Pareto inferior to some feasible option.

Consider, then, the following conditions:

> *Deontic Strong Pareto*: For any two alternatives x and y, if each person has at least as much benefits in x as in y, then, whenever both are feasible: (1) x is just, if y is, and (2) if there is at least one person that has more benefits in x than in y, then y is not just.

> *Deontic Weak Egalitarianism*: If two alternatives x and y are anonymously Pareto incomparable, then, whenever both are feasible: (1) if they are equally equal, one is just if and only if the other is, and (2) if x is more equal than y, then y is not just.

> *Deontic Moderate Egalitarianism*: If two alternatives x and y are Pareto incomparable, then, whenever both are feasible: (1) if they are equally equal, one is just if and only if the other is, and (2) if x is more equal than y, then y is not just.

Deontic Paretian Egalitarianism is the conjunction of Deontic Strong Pareto and Deontic Weak Egalitarianism. The conjunction of Deontic

---

[5] The axiological and deontic approaches to justice have been exhaustively studied in the social choice literature, where they are known as the social ordering and social choice function approaches. See, for example, Amartya Sen, "Social Choice Theory: A Re-Examination," *Econometrica*, XLV (1977): 53–89, reprinted in his *Choice, Welfare, and Measurement* (Cambridge: MIT, 1982), pp. 158–200.

Strong Pareto and Deontic Moderate Egalitarianism defines *Deontic Paretian Moderate Egalitarianism.* We shall starting by showing that, as in the axiological case, this latter theory of justice is impossible if one also accepts the following condition:

> *No Prohibition Dilemmas*: For any feasible set of alternatives, at least one feasible alternative is just.

This condition is plausible from the perspective of *practical justice,* according to which justice is purely a matter of comparing favorably in the relevant respects with the feasible alternatives (for example, being at least as good in the relevant respect as all (or 90%) of the feasible alternatives [which is always possible], as opposed to giving everyone an adequate level of benefits [which is not always possible]). Practical justice always satisfies No Prohibition Dilemmas. Of course, justice can be understood in a more ideal way that does not guarantee that at least one feasible alternative is just, but we here focus on practical justice.[6]

Before introducing our first deontic result, let us note that we are assuming (as is standard) that, with one qualification, any logically possible set of alternatives is a possible feasible set. The qualification is that we restrict feasible sets to those for which, for each individual, there is a maximum possible benefit. The qualification is needed because it is unclear what rationality and justice require when benefits can be increased without limit. Let us now note:

> *Result 4*: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas, and (3) the equality relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, then Deontic Paretian Moderate Egalitarianism is impossible.

The force of this result (and the structure of the proof) can be seen by considering the feasible set {<11,9,8>, <10,9,8>, <9,10,8>}. Note that <10,9,8> is more equal than <11,9,8> by Strong Contracting Extremes Equality, and hence Equality Strong Anonymity entails that <9,10,8> is also more equal. Deontic Paretian Moderate

---

[6] It is important to note that the choice function approach of social choice theory has No Prohibition Dilemmas implicitly built into it. This is because the concept of a choice function—which is a function that, for any given feasible set, selects the subset consisting of all and only the just alternatives relative to that feasible set—is defined to always select a nonempty set. We here posit No Prohibition Dilemmas for practical justice, but for an argument that prohibition dilemmas are conceptually possible (for example, not ruled out by deontic logic as such), see Vallentyne, "Prohibition Dilemmas and Deontic Logic," *Logique et Analyse,* XVIII (1987): 113–22, and "Two Types of Moral Dilemmas," *Erkenntnis,* XXX (1989): 301–18.

Egalitarianism thus makes the following judgments: (1) <11,9,8> is unjust (because it and <9,10,8> are Pareto incomparable and the latter is more equal), (2) <10,9,8> is unjust (by Deontic Strong Pareto), and (3) <9,10,8> is unjust because <10,9,8> is unjust and they are equally equal (by Deontic Moderate Egalitarianism). Thus, nothing is just, which violates No Prohibition Dilemmas.

Given the conditions of the result, Deontic Paretian Moderate Egalitarianism is impossible. It is not, however, the only form of Deontic Paretian Egalitarianism. Indeed, there are versions of Deontic Paretian Egalitarianism that, for some feasible sets, disagree with weak maximin. But this is not a particularly interesting observation, because, as we shall now show, if a feasibility set is rich in a certain sense, then the possibility of disagreement disappears.

Domain Richness guarantees that the set of all alternatives includes all logically possible distributions of benefits. This does not, however, guarantee that any particular feasibility set is rich in any interesting sense. We shall focus, therefore, on feasibility sets that are *minimally rich* in the sense that if alternative *x* is feasible and maximizes (relative to the feasibility set) the benefits to the worst off position, then so is an alternative that gives everyone those same benefits. Thus, the feasible set {<2,4,6> , <1,9,9>}, is not minimally rich because it does not contain <2,2,2>, whereas {<2,4,6> , <1,9,9>, <2,2,2>} is minimally rich.

We have the following result for feasible sets that are minimally rich:

> *Result 5*: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas, and (3) the equality relation satisfies Perfect Equality, then, *for any minimally rich set*, Deontic Paretian Egalitarianism (a) judges just some alternative that maximizes the benefits of the worst off position and (b) judges unjust all alternatives that do not maximize the benefits to the worst off position.

To see the force of this result, consider the minimally rich set {<2,4,6>, <3,3,6>, <3,3,3>}. Deontic Paretian Egalitarianism judges that (1) <2,4,6> is unjust (by Deontic Weak Egalitarianism, because it is less equal, by Perfect Equality, than the anonymously Pareto incomparable <3,3,3>), and (2) <3,3,3> is unjust (by Deontic Strong Pareto, because it is Pareto inferior to <3,3,6>). No Prohibition Dilemmas then entails that <3,3,6> is just. But <3,3,6> and <2,4,6> are anonymously Pareto incomparable, and thus it follows that Deontic Paretian Egalitarianism must judge <3,3,6> as more equal than <2,4,6> (since, if it did not, <2,4,6> would also be judged just, and that contradicts the first statement made in (1)). A suitable generalization of this shows that it is not possible for Deontic Paretian Egalitarianism to judge just some alternative that does not maximize the benefits to the worst off position, and has to judge just some alternative that does so if the set is minimally rich.

Thus, although Deontic Paretian Egalitarianism can disagree with weak maximin, it cannot do so for feasibility sets that are minimally rich. Given that minimal richness is a condition that is often satisfied in real life situations, this possibility of disagreement is not very significant. For a wide range of practical applications, Deontic Paretian Egalitarianism cannot contradict weak maximin.

It is worth emphasizing that this result does not depend on any assumption of Acyclicity or any similar deontic consistency condition. Below we shall establish some results that do so depend.

VI. DEONTIC JUSTICE RECONSIDERED

The results of the previous section establish that merely moving to a deontic framework is not enough to avoid the core of the two axiological results. One might argue, however, that the above formulations of the deontic egalitarian conditions are excessively strong. To see this, consider:

> *Restricted Deontic Weak Egalitarianism*: If, for a given feasible set, two alternatives *x* and *y* are anonymously Pareto optimal, then (1) if they are equally equal, one is just if and only if the other is, and (2) if *x* is more equal than *y*, then *y* is not just.
>
> *Restricted Deontic Moderate Egalitarianism*: If, for a given feasible set, two alternatives *x* and *y* are Pareto optimal, then (1) if they are equally equal, one is just if and only if the other is, and (2) if *x* is more equal than *y*, then *y* is not just.

These conditions are like their unrestricted counterparts, except that they apply, for a given feasible set, only to Pareto optimal alternatives (and, in the case of Restricted Deontic Weak Egalitarianism, only to those that are anonymously Pareto optimal)—rather than to all feasible alternatives. Suppose, for example, that the feasibility set is {<2,5>, <5,3>, <4,3>}. Deontic Weak Egalitarianism says that <2,5> is not just because it is anonymously Pareto incomparable to, and (by Equality Strong Anonymity) less equal than, <4,3>. Restricted Deontic Weak Egalitarianism, however, does not say this. It applies only to anonymously Pareto optimal options. Given that <4,3> is not anonymously Pareto optimal, the fact that it is more equal than <2,5> is not deemed relevant.

Restricted Deontic Paretian Egalitarianism is the conjunction of Deontic Strong Pareto and Restricted Deontic Weak Egalitarianism, and *Restricted Deontic Paretian Moderate Egalitarianism* is the conjunction of Deontic Strong Pareto and Restricted Deontic Moderate Egalitarianism.[7] Unlike the axiological and the unrestricted deontic case, in this restricted deontic case, Paretian Moderate Egalitarianism is, it seems, possible, as long as we assume that the equality relation satisfies the following condition:

> *Equality Acyclicity*: If, for alternatives $x_1,\ldots, x_n$, x_1 is more equal than x_2, x_2 is more equal than $x_3,\ldots$, and x_{n-1} is more equal than x_n, then x_n is not more equal than x_1.

We have been assuming that the justice relation is acyclic, but so far we have not needed to assume that the equality relation is. Without this assumption, every Pareto optimal alternative could be less equal than some other Pareto optimal alternative, and Restricted Deontic Paretian Moderate Egalitarianism would judge all feasible alternatives unjust, which would violate No Prohibition Dilemmas. The assumption of acyclicity, however, is entirely uncontroversial, and we are merely explicitly noting it.

Consider then:

> *Result 6*: Given (1) Benefitism and Domain Richness and (2) that the equality relation satisfies Equality Acyclicity, Perfect Equality, Equality Strong Anonymity, Strong Contracting Extremes Equality, there are versions of Restricted Deontic Paretian Moderate Egalitarianism (and hence Restricted Deontic Paretian Egalitarianism in general) that satisfy No Prohibition Dilemmas and that, even for a minimally rich feasible set, (a) judge a feasible alternative just even though it does not maximize the benefits to the worst off position, and (b) judge a feasible alternative unjust even though it does maximize the benefits to the worst off position.

To see this, consider the following version:

> *Full Restricted Deontic Paretian Moderate Egalitarianism*: An alternative is just, relative to a feasible set, if and only if it is Pareto optimal and no Pareto optimal alternative is more equal.[8]

This theory judges all alternatives just except those that are not Pareto optimal and those that are less equal than some Pareto optimal alternative. Hence, it satisfies both Deontic Strong Pareto and Restricted Deontic Moderate Egalitarianism. Moreover, for any acyclic equality relation, it will also satisfy No Prohibition Dilemmas (since there will always be at least one Pareto optimal alternative that is not less equal than some other one). Hence, in order to see that this theory of justice constitutes a real alternative to weak maximin, we may simply consider an equality relation satisfying our minimal conditions that disagrees with weak maximin. By way of illustration, suppose that a distribution is more equal if the total shortfall from the mean is less. This equality relation is acyclic and satisfies our three conditions on equality (proof omitted). Now consider the feasible set {<2,2,2>, <2,3,52>, <1,40,46>}, which is minimally rich. Alternatives <2,3,52> and <1,40,46> are the only two Pareto optimal alternatives, and <2,3,52> (mean of 19, total shortfall of 33) is less equal than <1,40,46> (mean of 29, total shortfall of 28). Hence, given this equality relation, Full Restricted Deontic Paretian Moderate Egalitarianism judges <1,40,46> as just (because it is a most equal Pareto optimal alternative), judges <2,3,52> as unjust (because it is less equal than some other Pareto optimal alternative than), and judges <2,2,2> as unjust (because it is not Pareto optimal). Thus, Restricted Deontic Paretian Moderate Egalitarianism can judge unjust an alternative that maximizes the benefits to the worst off position, and it can judge just an alternative that does not maximize these benefits.

---

[7] We have not formulated a restricted version Deontic Strong Pareto because it is equivalent to the unrestricted version.

[8] Roughly this view is defended in Rex Martin, *Rawls and Rights* (Lawrence: Kansas UP, 1985), chapter 6.

Finally, then, we have, it seems, the possibility of a form of Paretian egalitarianism that can disagree with weak maximin even in the case of minimal rich feasible sets. There may, however, be further plausible conditions that, if imposed, would rule out this apparent possibility. Before concluding, we shall consider one such candidate condition.

As noted above, we have not assumed any consistency conditions that correspond to Acyclicity (or related conditions). In a deontic context, however, the following condition is widely accepted:

> *Alpha* (contraction consistency): If an alternative is judged just relative to a given feasible set, then it is also judged just from any subset containing it.

Alpha is concerned with the preservation of judgment of justness as a feasibility set is contracted while still containing the alternative originally judged just. It is violated, for example, if x is judged just relative to $\{x,y,z\}$, but not judged just relative to $\{x,y\}$. If it is a winner against x and y, why would not it also be a winner against y alone?

If we accept Alpha, the impossibility of Paretian Moderate Egalitarianism returns:

> *Result 7*: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas and Alpha, and (3) the equality relation satisfies Equality Strong Anonymity and Strong Contracting

Extremes Equality, then Restricted Deontic Paretian Moderate Egalitarianism is impossible.[9]

To see the force of this result, consider the feasible set consisting of <6,9,12>, <11,9, 6> and <100,9,6>. Note that <6,9,12> is more equal than <100,9,6> (by Strong Contracting Extremes Equality and Equality Strong Anonymity). Restricted Deontic Paretian Moderate Egalitarianism thus makes the following judgments: (1) <11,9,6> is not just (by Deontic Strong Pareto), (2) <100,9,6> is not just (by Restricted Deontic Moderate Egalitarianism, because it is less equal than the Pareto optimal <6,9,12>). No Prohibition Dilemmas then entails that <6,9,12> is just. We shall now show that this leads to a violation of Alpha by considering the feasible set {<6,9,12>, <11,9,6>}, a subset of the original set. Note that <6,9,12> is less equal than <11,9,6> (by Strong Contracting Extremes Equality and Equality Strong Anonymity). Restricted Deontic Moderate Egalitarianism here judges <6,9,12> as not just (by Restricted Deontic Moderate Egalitarianism). Thus, given that <6,9,12> was judged just from the first set, but not from this proper subset, Alpha is violated.

If we impose Alpha, then Restricted Deontic Paretian Moderate Egalitarianism is impossible. It is not, however, the only form of Restricted Deontic Paretian Egalitarianism and some of these other forms are possible. As in the previous cases we have considered, however, no form can disagree with weak maximin when applied to minimally rich sets.

> *Result 8*: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas and Alpha, and (3) the equality relation satisfies Perfect Equality, then, for any feasible set, Restricted Deontic Paretian Egalitarianism must (a) judge just some alternative that maximizes the benefits to the worst off position, and (b), *if the feasible set is minimally rich,* judge unjust any alternative that does not maximize the benefits to the worst off position.

To see the force of this result, consider the minimally rich feasible set <2,4,6>, <3,3,3>, <3,3,6>. Restricted Deontic Paretian Egalitarianism judges <3,3,3> as unjust (by Deontic Strong Pareto). To see how it judges <2,4,6>, consider the following subset of the initial set, {<2,4,6>, <3,3,3>}. In this case, Restricted Deontic Paretian Egalitarianism judges <2,4,6> as unjust (by Restricted Deontic Weak Egalitarianism, because it is less equal than the anonymously Pareto optimal <3,3,3>). Alpha then entails that <2,4,6> is also unjust relative to the expanded set. No Prohibition Dilemmas then entails that <3,3,6> is just relative to this expanded set. The proof generalizes this example to all minimally rich feasible sets. Notice moreover that, because <2,4,6> and <3,3,6> are each anonymously Pareto optimal, Restricted Deontic Paretian Egalitarianism must hold that <3,3,6> is more equal than <2,4,6>. More generally, a distribution that is maximin better cannot be judged less equal by this theory. There is, that is, no possibility of measuring equality in a way that significantly diverges from weak maximin.

---

[9] The first to note the tension between Paretian egalitarianism and Acyclicity in the axiological case, and Alpha in the deontic case, was perhaps Albert Weale, "The Impossibility of Liberal Egalitarianism," *Analysis*, XL (1980): 13–19. The same point was made independently by Andrew Williams, "The Revisionist Difference Principle," *Canadian Journal of Philosophy*, XXV (1995): 257–81. Each wrongly concludes, however, that Paretian egalitarianism is incompatible with these conditions, whereas the correct conclusion is that it is incompatible if it measures equality in ways that disagree with weak maximin.

Do these results show that even Restricted Deontic Paretian Egalitarianism must agree with weak maximin? That, of course, depends on the plausibility of Alpha. One of us is firmly inclined to reject it. We shall end with a discussion of the strengths and weaknesses of this condition.

If axiological justice is a complete ordering, and, for any given feasible set, the set of just alternatives is simply the set of best feasible alternatives, then Alpha will be satisfied. Alpha is clearly a desirable condition for justice in the sense that it would be nice if justice satisfied it. The question, however, is whether it is a *mandatory* condition—one that any minimally adequate conception of justice *must* satisfy.

Restricted Deontic Paretian Moderate Egalitarianism can violate Alpha because it requires that just options be no less good in some dimension (equality) than any alternative that is Pareto optimal. Removing an option can make an originally Pareto suboptimal option become Pareto optimal, and this option may be better (more equal) than some third option that was originally not less good than any Pareto optimal option. Thus, this third option may be just relative to the original set but not relative to the subset. Suppose for the sake of argument that <2,2,2> is more equal than <1,40,46>, and that the latter is more equal than <2,3,43>. In this case, Restricted Deontic Paretian Moderate Egalitarianism judges just <1,40,46> when the feasible alternatives are <2,2,2> and <2,3,43> (since it is Pareto optimal and more equal than <2,3,43> and <2,2,2> is not Pareto optimal), but it judges <1,40,46> unjust when the only feasible alternative is <2,2,2> (since the former is less equal and both are Pareto optimal). Why would this be incompatible with minimum adequacy in a theory of justice? It cannot be simply because there are two distinct dimensions of value. Justice can surely be pluralistic. Nor can it be the lexical priority given to one dimension (Pareto optimality). Minimally adequate theories of justice can surely invoke lexical priority. Nor, it seems, can it be that the lexically prior dimension has the feature that removing an option can make admissible a previously inadmissible option in this dimension. Thus, it is unclear why Alpha would be a mandatory condition on justice.

Several authors have argued against Alpha (or its counterparts formulated in terms of rationality and the like rather than justice). Amartya Sen, for example, has argued against any kind a priori requirement of internal consistency, and against Alpha in particular, for rational individual choice or social choice.[10] There is nothing irrational, he argues, about having the goal of choosing the second largest piece of cake (for example, so as not to appear greedy). This violates Alpha, however, since the originally second largest piece will not be rationally chosen if the largest piece is removed as a possible choice. Likewise, Sen argues, there is nothing irrational about choosing to fast when one has the alternatives of a full meal and of a very limited meal, but not choosing to fast when one's only alternative is the very limited meal. The symbolic significance of fasting is radically altered by the absence of the possibility of full meal.

Moreover, there are many examples of conceptions of justice that violate Alpha, but are not wildly implausible. To start, suppose that an option is judged just, relative to a feasible set, if and only if it is in the top 10% of the feasible alternatives in terms of some complete order of the value of alternatives. This is a kind of satisficing theory. Consider an alternative that is barely in the top 10% and then remove many of the less valuable options. If enough are removed, the original alternative will cease to be in the top 10%, and thus cease to be just. This violation of Alpha does not, however, seem wildly implausible. It is not clearly inappropriate for justice to take into account the values of other feasible alternatives and to judge just those that are near the top in relative terms.

Consider next the bargaining solution (and view of justice), advocated by David Gauthier.[11] It requires that "relative benefit" be maximinned, where relative benefit for a given individual is the ratio of the benefit the individual receives to the maximum feasible benefit she could receive. Consider the feasible set {<3,1>, <2,4>, <10,0>}. The relative benefits of these three alternatives are respectively {< .33,.25>, <.2,1>,

---

[10] See "Internal Consistency of Choice," *Econometrica*, LXI (1993): 495–521; reprinted in Sen's *Rationality and Freedom* (Cambridge: Harvard, 2002), pp. 121–57.

[11] See *Morals by Agreement* (New York: Oxford, 1986).

<1,0>}, and thus <3,1> alone is deemed just (since .25 is greater than .2 and 0). If <10,0> is removed, however, the relative benefits are respectively {<1,.25>, <.66,1>}, and thus <2,4> alone is judged just (since .66 is greater than .25). This violation of Alpha, however, does not seem wildly implausible. Moreover, it occurs even if the requirement to maximin relative benefits is replaced with the requirement to maximize the total relative benefits or the requirement to maximize some kind of prioritarian weighted total relative benefit. It is not clearly inappropriate to scale benefits relative to the most an individual could get in a situation.

Consider next the view of justice that requires leximaxing complaints (that is, minimizing the maximum complaint, and in case of ties minimizing the second (third and so on) maximum complaint), where a person's complaint is the absolute shortfall from the greatest benefits she could have gotten in the given choice situations. (This is the analogue of maximin regret in individual choice under uncertainty.) Consider the feasible set {<4,6>, <5,4>, <10,0>}. The individual complaints for these three alternatives are respectively {<6,0>, <5,2>, <0,6>}, and so <5,4> alone is just (since 5 is less than 6). If <10,0> is removed, however, the individual complaints are respectively {<1,0>, <0,2>}, and so <4,6> alone is judged just. Again, this violation of Alpha does not seem wildly implausible. Moreover, as above, this violation occurs even if the requirement to leximax complaints is replaced with the requirement to minimize the total complaints or the requirement to minimize some kind of prioritarian weighted total complaint. It is not clearly inappropriate to have justice depend on complaints in these ways.

Finally, consider a theory of justice that Isaac Levi views as not implausible.[12] It holds that (1) the weights assigned to each person's benefits are nonnegative and sum to one, (2) there may be more than one admissible way to weigh each person's benefits (perfectly equal weighting may not be required), and (3) an option is just if and only if it leximins benefits relative to the feasible options that maximize average benefits according to *at least one* admissible weighting scheme. Consider the feasible set consisting of <9,1>, <1,9>, and <2,2>. Option <9,1> maximizes average benefits when at least 50% of the weight is given to the first person, and <1,9> does so when at least 50% is given to the second person. Option <2,2>, however, does not maximize average benefits on any admissible weighting scheme. Thus, <9,1> and <1,9> are each just relative to this set, since each leximins benefits relative to the options that maximize average benefits relative to at least one weighting scheme. Consider now the feasible set consisting of just <9,1> and <2,2>. Here again <9,1> maximizes average benefits relative to some weighting schemes. In this case, however, so does <2,2> (for example, for any weighting scheme that gives the second person at least 7/8 of the weight). Only <2,2>, however, leximins relative to the options that maximize the average on some weighting. Hence, <9,1> is not just relative to this subset, and that violates Alpha. Allowing some such indeterminacy in weightings (for example, to allow for up to some specified extra weight to given to the chooser) is not clearly implausible, nor is Levi's approach for dealing with such situations.[13]

---

[12] See *Hard Choices* (New York: Cambridge, 1986), chapter 9.6. Levi also defends violations of alpha for individual rational choice where there are indeterminate probabilities (chapter 7.9) and where there is ordinal indeterminacy in the value for an agent (chapter 6.9).

In sum, if an alternative is just if and only if it is a best feasible alternative, where the goodness of alternatives is a complete ordering, then Alpha will be satisfied. Alpha need not be satisfied, however, if deontic justice is not *rationalizable* in the sense of there being some (context-free) ranking relation *R*, over alternatives, such that an alternative is just if and only if it is an *R*-best feasible alternative. Although rationalizability is a nice property, there seems to be little reason to require it of all minimally adequate conceptions of justice. As we have illustrated above, deontic justice may rely on normative concepts that are not compatible with a context-free ranking relation.

We therefore tentatively conclude that Restricted Deontic Paretian Egalitarianism is possible and can disagree with weak maximin.

VII. CONCLUSION

To be plausible, egalitarian theories of justice cannot be pure egalitarian theories. Axiological versions (what is at least as just as what) must hold that (1) Pareto improvements (making some individuals better off and no one worse off) make things more just and deontic versions (what is just relative to the set of feasible alternatives) must hold that no Pareto suboptimal alternative is just. Pareto efficiency, that is, is lexically prior to the demands of equality. Paretian egalitarianism is committed to this view, and we have explored some of the forms that it can take.

---

[13] For further criticisms of Alpha, see Edward F. McClennen, *Rationality and Dynamic Choice* (New York: Cambridge, 1990), chapter 2; Piers Rawling, "Expected Utility, Ordering, and Context Freedom," *Economics and Philosophy*, XIII (1997): 79–86; and Paul Weirich, *Equilibrium and Rationality: Game Theory Revised by Decision Rules* (New York: Cambridge, 1998), chapter 4.

Given the various plausible conditions that we identified, we have established that axiological Paretian egalitarianism (1) is impossible—if it makes equality dominant in all cases of Pareto incomparability, and (2) cannot disagree with weak maximin, if it limits the role of equality to cases of anonymous Pareto incomparability. It cannot, for example, hold that <2,10,100> is less just than <1,100,100>. Indeed, given that these two alternatives are anonymously Pareto incomparable (and thus ranked on the basis of equality), it cannot hold the former is less *equal* than the latter. This radically limits the admissible measures of equality available to Paretian Egalitarianism. Indeed, none of the standard measures of equality are available, since they all hold that at least sometimes a maximin worse distribution is more equal. (For example, above, the Gini coefficient of inequality is 0.58 for the first and 0.33 for the latter alternative.) Axiological Paretian egalitarianism thus seems quite unpromising as an alternative to leximin and other forms of weak maximin.

If we turn to deontic justice, roughly these results remain valid if the egalitarian conditions are taken to require that an alternative be judged unjust if some *feasible* alternative is more equal. If, however, the egalitarian requirements are restricted to require only that an alternative be judged unjust if some *Pareto optimal* alternative is more equal, then Paretian egalitarianism can disagree with weak maximin. This is the only promising form of Paretian egalitarianism. Even it, however, cannot disagree with weak maximin if Alpha (contraction consistency) is imposed. We have suggested that Alpha is not a mandatory condition for justice, but admittedly our discussion is not conclusive. The future of Paretian egalitarianism depends on this issue and thus further investigation is needed.

BERTIL TUNGODDEN
Norwegian School of Economics and Business Administration and
Chr. Michelsen Institute

PETER VALLENTYNE
University of Missouri-Columbia

APPENDIX

Result 1: If (1) we have Benefitism and Domain Richness, (2) the equality relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, and (3) there are only two people, then for any two anonymously Pareto incomparable alternatives, the alternative that makes the worst off position better off is more equal.

Proof: Consider any two anonymously Pareto incomparable alternatives. Because there are only two persons and they are anonymously

Pareto incomparable, if the worst off position in one alternative, *x*, receives more benefits then the worst off position in the other alternative, *y*, then the best off position in *x* receives fewer benefits than the best off position in *y*. If the worst off person is the same in both alternatives (for example, *x*=<6,8> and *y*=<5,100>), then it follows directly from Strong Contracting Extremes Equality that *x* is more equal than *y*. If the worst off person is not the same in both alternatives (for example, *x*=<6,8> and *y*=<100,5>), then by Domain Richness, we may consider a permutation of *x*, *x** (*x**= <8,6>), and Strong Contracting Extremes Equality will judge *x** more equal than *y*. Hence, by Equal Strong Anonymity, *x* is more equal than *y*.

> *Result 2*: If (1) we have Benefitism and Domain Richness, (2) the justice relation satisfies Acyclicity, and (3) the equality relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, then Paretian Moderate Egalitarianism is impossible.

Proof: By Domain Richness, we may consider two alternatives *x* and *y* such that (a) they are Pareto incomparable but not anonymously so, (b) the benefits are the same in every position except for the best off position, and (c) the benefits in the best off position are higher in *x* than *y* (for example, *x* = <7,8,11> and *y* = <10,8,7>). (1) By Domain Richness, there is an alternative, *y** (<7,8,10>) that is a permutation of *y* and such that Strong Contracting Extremes Equality entails that *y** is more equal than *x*. Equality Strong Anonymity then entails that *y* is more equal than *x*. Hence, Weak Egalitarianism judges that *y* is more just than *x* (because they are Pareto incomparable and *y* is more equal). (2) A contradictory judgment, however, can be obtained as follows. By Domain Richness, there is an alternative *z* (for example, <7,8,9>) such that (a) *x* is Pareto superior to *z*, (b) Strong Contracting Extremes Equality judges a permutation of *z*, *z** (for example, <9,8,7>), more equal than *y*, and (c) *z* and *y* are Pareto incomparable. Paretian Moderate Egalitarianism thus makes the following judgments: *x* is more just than *z* (by Strong Pareto), and *z* is more just than *y* (*z* is more equal than *y* by Equality Strong Anonymity, and thus more just than *y* by Weak Egalitarianism). By Acyclicity, we have that *y* is *not* more just than *x*, which contradicts the conclusion obtained in (1).

> *Result 3*: If (1) we have Benefitism and Domain Richness, (2) the justice relation satisfies Acyclicity, and (3) the equality relation satisfies Perfect Equality, then Paretian Egalitarianism judges that a distribution is *not less just* than another if it gives greater benefits to the worst off position. Moreover, (a) if the justice relation also satisfies Consistency, then Paretian Egalitarianism judges that a distribution is *not equally or less just*, if it gives

> greater benefits to the worst off position, and (b) if the justice relation also satisfies Quasi-Transitivity, then Paretian Egalitarianism judges that a distribution is *more just* than another if it gives greater benefits to the worst off position.

Proof: Consider any x and y, where the worst off position in x gets more benefits than the worst off position in y. (1) If no one is worse off in x than in y, then it follows from Strong Pareto that x is more just than y, as required. (2) If someone is worse off in x than in y, then the two are Pareto incomparable, and we consider two cases: (a) If there is perfect equality in x, then it follows that x and y are anonymously Pareto incomparable (for example, $x = <3,3,3>$ and $y = <2,4,10>$), and that x is more equal. Weak Egalitarianism then entails that x is more just than y, as required. (b) If there is inequality in x (for example, $x = <3,3,5>$ and $y = <2,4,10>$, then by Domain Richness there exists an alternative, z (for example, <3,3,3>) such that everyone in z gets the same as the worst off in x. By Strong Pareto, x is more just than z. In this case, z and y are anonymously Pareto incomparable and Perfect Equality entails that z is more equal than y. Weak Egalitarianism then judges z more just than y. We thus have that x (<3,3,5>) is more just than z (<3,3,3>), which is more just than y (<2,4,10>). Acyclicity then entails that x is not less just than y, Consistency entails that x is not equally or less just than y, and Quasi-Transitivity entails that x is more just than y, as required.

> *Result 4*: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas, and (3) the equality relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, then Deontic Paretian Moderate Egalitarianism is impossible.

Proof: We will illustrate the proof with a three person example: by Domain Richness, we may consider the following feasible set {<11,9,8>, <10,9,8>, <9,10,8>}. Here, <10,9,8> is more equal than <11,9,8> by Strong Contracting Extremes Equality, and hence Equality Strong Anonymity entails that <9,10,8> is also more equal. Deontic Paretian Moderate Egalitarianism thus makes the following judgments: (1) by Deontic Weak Egalitarianism, <11,9,8> is unjust because it and <9,10,8> are Pareto incomparable and the latter is more equal, (2) by Deontic Strong Pareto, <10,9,8> is unjust because it is not Pareto optimal, and (3) by Deontic Weak Egalitarianism, <9,10,8> is unjust because <10,9,8> is unjust and they are equally equal (by Equal Strong Anonymity). Thus, nothing is just, which violates No Prohibition Dilemmas.

> *Result 5*: If (1) we have Benefitism and Domain Richness, (2) deontic

justice satisfies No Prohibition Dilemmas, and (3) the equality relation satisfies Perfect Equality, then, *for any minimally rich set*, Deontic Paretian Egalitarianism (a) judges just some alternative that maximizes the benefits of the worst off position and (b) judges unjust all alternatives that do not maximize the benefits to the worst off position.

Proof: Consider any minimally rich feasible set, and let x be any feasible alternative that does not maximize the benefits to the worst off position and let y be any feasible alternative that does maximize the benefits to the worst off position. We will now show that Deontic Paretian Egalitarianism has to judge x as unjust in this situation and that it must judge just some alternative that maximizes the benefits to the worst off position. (1) If no else gets more benefits in x than in y (for example, $x = <2,4,6>$ and $y = <3,4,6>$), then Deontic Strong Pareto entails that x is unjust, as required. (2) If someone else gets more benefits in x than in y (for example, $x = <2,4,6>$ and $y = <3,3,6>$), then it follows that there is inequality in x. In this case, given that the feasible set is minimally rich, there exists an alternative, z (for example, $<3,3,3>$), such that everyone in z gets the same benefits as the worst off position in y. Given that (a) alternative z is anonymously Pareto incomparable with x (since the benefits of the worst off position are higher in z than x, and the benefits of the best off position are lower in z than in x), and (b) Perfect Equality entails that z is more equal than x, Deontic Weak Egalitarianism entails that x is not just relative to this feasible set, as required. Given that x was an arbitrary feasible alternative that does not maximize the benefits to the worst off position, it follows that no such alternative is just. (3) By No Prohibition Dilemma, it follows immediately that some feasible alternative that does maximize the benefits to the worst off position is just, as required.

Result 6: Given (1) Benefitism and Domain Richness and (2) that the equality relation satisfies Equality Acyclicity, Perfect Equality, Equality Strong Anonymity, Strong Contracting Extremes Equality, there are versions of Restricted Deontic Paretian Moderate Egalitarianism (and hence Restricted Deontic Paretian Egalitarianism in general) that satisfy No Prohibition Dilemmas and that, even for a minimally rich feasible set, (a) judge a feasible alternative just even though it does not maximize the benefits to the worst off position, and (b) judge a feasible alternative unjust even though it does maximize the benefits to the worst off position.

Proof: Given in text.

Result 7: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas and Alpha, and (3) the equality

> relation satisfies Equality Strong Anonymity and Strong Contracting Extremes Equality, then Restricted Deontic Paretian Moderate Egalitarianism is impossible.

Proof: (1) By Domain Richness, we may consider the feasible set consisting of <6,9,12>, <11,9,6> and <100,9,6>. Here <6,9,12> and <100,9,6> are Pareto optimal, but <11,9,6> is not. (a) By Deontic Strong Pareto, <11,9,6> is not just. (b) By Equality Strong Anonymity and Strong Contracting Extremes Equality, <6,9,12> is more equal than <100,9,6>. Restricted Deontic Moderate Egalitarianism thus judges <100,9,6> not just. (c) By No Prohibition Dilemmas, <6,9,12> is thus just. (2) Consider now the feasible set consisting of just <6,9,12> and <11,9,6>, a subset of the original set. (a) By Equality Strong Anonymity and Strong Contracting Extremes Equality, <6,9,12> is less equal than <11,9,6>. Restricted Deontic Moderate Egalitarianism, thus judges <6,9,12> not just. (3) Given that <6,9,12> was judged just from the first set (in 1c), but not from this proper subset, Alpha is violated.

> *Result 8*: If (1) we have Benefitism and Domain Richness, (2) deontic justice satisfies No Prohibition Dilemmas and Alpha, and (3) the equality relation satisfies Perfect Equality, then, for any feasible set, Restricted Deontic Paretian Egalitarianism must (a) judge just some alternative that maximizes the benefits to the worst off position, and (b), *if the feasible set is minimally rich,* judge unjust any alternative that does not maximize the benefits to the worst off position.

Proof: Consider any feasible set for which some but not all alternatives maximize the benefits to the worst off position. Let *y* and *x* be arbitrary alternatives in this set that respectively maximize, and do not maximize, the benefits to the worst off position. We will first show that, if the feasible set is minimally rich, Restricted Deontic Paretian Egalitarianism has to judge *x* as unjust and has to judge just some alternative that maximizes the benefits of the worst off position. Following that, we show that it must make the second judgment even if the feasible set is not minimally rich. (1) Suppose, then, that the feasible set is minimally rich. (a) If no else gets more benefits in *x* than in *y* (for example, *x* = <2,4,6> and *y* = <3,4,6>), then Deontic Strong Pareto entails that *x* is unjust. Given that *x* was an arbitrary alternative that does not maximize the benefits to the worst off position, No Prohibition Dilemmas then implies that some alternative that does maximize the benefits to worst off position is just. (b) If someone else gets more benefits in *x* than in *y* (for example, *x* = <2,4,6> and *y* = <3,3,6>), then it follows that there is inequality in *x*. In this case, given that the feasible set is minimally rich, there exists an alternative, *z* (for example,

<3,3,3>), such that everyone in *z* gets the same benefits as the worst off position in *y*. Consider now the feasible set obtained from the original by removing all the alternatives other than *z* that maximize the benefits to the worst off position. In this restricted set, *z* is anonymously Pareto optimal and, by Perfect Equality, more equal than any other anonymously Pareto optimal alternative. Hence, by Restricted Deontic Weak Egalitarianism, *x* and all other alternatives except for *z* are unjust, and by No Prohibition Dilemmas *z* is just. By Alpha, *x* is also unjust in the original superset. Given that *x* was an arbitrary alternative that did not maximize the benefits to the worst off position, this shows that every alternative that fails to so maximize is unjust in the original set. No Prohibition Dilemmas then entails that some alternative that does maximize the benefits to the worst off position is just, as required. (2) Suppose now that the original set is not minimally rich. We may make this set minimally rich by adding an alternative that gives everyone the same benefits as the worst off position in an alternative that maximizes the benefits of the worst off. From (1), we know that for a minimally rich feasible set, there is an alternative, *y*, that maximizes the benefits of the worst off position, and is judged just, relative to this extended set. By Deontic Strong Pareto, it follows that *y* cannot be the alternative added to the initial set. Hence, *y* is in the initial feasible set. By Alpha, it follows that *y* is also just in the initial not minimally rich set, as required.

BOOK REVIEWS

The Realm of Reason. CHRISTOPHER PEACOCKE. New York: Oxford University Press, 2004. x + 284 p. Cloth $35.00*

This ambitious work aims to set out a "moderate rationalist" position in epistemology and ethics. Section I below provides noncritical exposition. Section II registers criticisms.

I. EXPOSITION

Peacocke argues for a "generalized rationalism," holding that "all entitlement has a fundamentally a priori component" (2). But his rationalism "differs from those of Frege and Gödel, just as theirs differ from that of Leibniz." He requires both substantive theories of intentional content and of understanding, and systematic formal theories of referential semantics and truth. We need an externalist theory of content: "Only mental states with externally individuated contents can make judgements about the external, mind-independent world rational" (123).

Purely evidential conceptions of meaning and content are inadequate (34–49). They cannot account for the following: a thinker often has to work out what would be evidence for a content; contents cannot depend, for their identity, on all of the infinitely ramifying evidential connections among them; and thinkers conceive, however tacitly, of (at least some) observed properties as categorical. By contrast with an evidential theory, a truth-conditional theory of content can account for all these problematic facts.

Peacocke states, develops, and defends three principles of rationalism which collectively "relate entitlement to truth, to the identity of states and their intentional contents, and to the a priori" (3–4). He does not thoroughly explain his central notion of entitlement, but this much is clear: any thinker is entitled to various transitions *in*, or *into*, thought. An example of a transition *into* thought would be that from one's perceptual experience to an observational judgment. An example of a transition *in* thought would be a logical inference from certain premises to a conclusion.

---

* I would like to thank Jon Cogburn, Pietro Giannone, Robert Kraut, William Melanson, Diana Raffman, William Roche, and Joseph Salerno for their helpful comments on earlier drafts. I am solely responsible for the contents of this review, including any errors that remain.

0022-362X/05/0203/155–62

A transition is rational just in case the thinker is entitled to it. (Note that this aims to explain rationality in terms of entitlement, not the other way round.) It is clear from p. 28 that Peacocke needs an abstract ontology of entitlements (such as proofs, in the case of mathematics). Yet he does not endorse "Gödel's obscure quasi-perceptual and quasi-causal epistemology of mathematics and the abstract sciences" (54).

Underlying any thinker's entitlements are objective norms (7). Peacocke's conception of entitlement is neither conventionalist nor language-relative (51). Attitudes and community practices do not suffice to institute norms (57). There are three levels of entitlement: instances; generalization about instances; and explanations of those generalizations (60–65). The third level, on which Peacocke places special emphasis, is where we encounter the brunt of the philosophical work to be done by the moderate rationalist.

To show that a transition is rational, we need to allude to reference and truth (15). If a transition tends to lead to, or to preserve, truth, let us say that it *conduces* to truth. Peacocke's three principles (on pp. 11, 52, and 148) boil down to:

(I) Rational transitions are truth-conducive.
(II) The truth-conduciveness of rational transitions is to be explained by the intentional contents and states involved.
(III) Rational transitions are a priori.

One must not underestimate the reach of the a priori. "[F]or any concept at all there will be a priori principles governing it" (193); and "[a]ny case of knowledge of an empirical theory exists only because some a priori entitlements also exist" (194). Any empirical theory involves the a priori (195) in at least three ways: in its methodological canons of confirmation, induction, and abduction; in our taking perceptual experience and memory at face value; and in our employing at least some logical principles (and, one might add: mathematics).

Peacocke defines perceptions as *instance-individuated* with respect to their contents just in case

> [w]hat makes these perceptions have the content they do is the fact that *when the subject is properly related to the world* the holding of these contents causally explains such perceptual experiences of their holding (69, emphasis added).

We are entitled, Peacocke says, to take experiences with instance-individuated contents at face value. He asks "What is the philosophical explanation of the existence of this entitlement" (74)? After a digression on complexity-reducing explanations, he ventures the suggestion that:

> the explanation of the occurrence of a perceptual experience with the instance-individuated content that *p* which most reduces complexity is that the experience is produced by a device which has evolved by natural selection to represent the world to the subject (87).

Chapter 5 is devoted to the rationality of enumerative induction: those transitions from finite collections of evidence $Fa_1 \& Ga_1$, ..., $Fa_n \& Ga_n$ to the conclusion 'All *F*s are *G*'. We have, says Peacocke, an a priori, though defeasible, entitlement to make such transitions.

Basic moral principles, according to Peacocke, are *contentually a priori*. (I spare the reader here the contorted definition of this technical term. It can be found on p. 160.) Our ability to know them traces back to our grasp of moral concepts. The notion of truth that applies to moral propositions is uniform with that which applies in any other domain (233). The moral rationalist turns out to be a moral realist also, in holding a species of mind-independence about moral truth (234).

II. CRITICISMS

The writing is often indigestible and obscure. It imparts an appearance of complex depth to ideas that are often simple, superficial, or passé. Some of Peacocke's thoughts may be profound, but his writing invites the charge of unnecessary complication. It can garb thoughts so fancily that it can take a few moments' reflection to realize that they are obvious.

Peacocke draws on his earlier work *A Study of Concepts*.[1] In so doing, however, he repeats (187–88) his mistake, in that work, of thinking that one can quantify into positions occupied by numerical adjectives rather than substantivals.[2]

The impression to be gained, as this work begins, is of an a priorist forsaking his armchair to greet those for whom philosophy has long been taken as being continuous with science. The blinding by science begins in chapter 3, with snowflakes. After a brief flurry, a thaw sets in. By chapter 7 science is once again ignored. One is back in one's soggy armchair for the discussion of moral rationalism, deprived of any insights to be had from neuroscience, social psychology, or sociobiology.

---

[1] Cambridge: MIT, 1995.

[2] For a detailed discussion of this shortcoming, and others, in *A Study of Concepts*, see my "The Emperor's New Concepts," *Philosophical Perspectives*, XVI (2002): 345–77.

Any reader hostile to the a priori will ask: What is the point of talking about the a priori? Why not suppress 'a priori' wherever it occurs? One will still have that a transition is rational just in case the thinker is entitled to it. The three levels of entitlement could still be recognized, and accorded due theoretical interest. One would still have principles (I) and (II). Only Principle (III) would have to go, to be replaced, perhaps, with something like:

> (III*) Rational transitions are highly entrenched.

By 'highly entrenched' here I mean so positioned and so influentially at work within the individual thinker's web of belief as to be relinquished only with the utmost reluctance. Any empirical theory will involve the highly entrenched in the same three ways as Peacocke claimed it would involve the a priori: in its methodological canons, in the primacy accorded reliable observations and memories, and in the principles of logic and mathematics used. Finally, one could still insist that to show that a transition is rational, we need to allude to reference and truth.

Peacocke has no interest in whether there are any synthetic a priori truths, and, if so, how this is possible. Hilary Putnam, for one, viewed W.V. Quine's celebrated attack in "Two Dogmas of Empiricism" on the analytic/synthetic distinction as being just as devastating for the a priori/a posteriori distinction (and held that that was how Quine saw it too).[3] Nevertheless, Peacocke abandons the former distinction, in what he thinks is due deference to Quine, but takes himself as still entitled to the latter distinction.

Peacocke defines a content as a priori

> if a thinker can be entitled to accept it without the entitlement being constitutively dependent upon the content or kind of her perceptual experiences *or other conscious states* (24–25, emphasis added).

But what about reflective awareness of the content itself? On a reading of this definition as a one-way conditional, '*Cogito, ergo sum*' would not yet be confirmed as enjoying a priori status; and on a biconditional reading, the *Cogito* would be denied a priori status. Perhaps Peacocke could avoid this problem by re-classifying it as *judgmentally a priori with respect to reflection* (see p. 160). Or perhaps one ought to say that the claim that I exist is *judgmentally a priori with respect to my thinking.* There is no guidance from Peacocke as to how one should classify the *Cogito* within his own rationalist scheme of things, and using his special terminology.

---

[3] Quine, "Two Dogmas of Empiricism," in *From a Logical Point of* View (Cambridge: Harvard, 1961, rev. ed.), pp. 20–46; Putnam, "There Is at Least One *A Priori* Truth," in *Realism and Reason: Philosophical Papers, Volume III* (New York: Cambridge, 1983), pp. 98–114 (see the opening paragraph).

Peacocke also significantly misunderstands the structure of his own argument in response to skepticism. He is aware of the charge that, by invoking explanations from the theory of natural selection, he will be resting his account of the a priori on empirical considerations. He resists the charge (97–98). He claims that his a priori explanation has taken the form of the following argument:

> [Perceptual e]xperiences are complex events.
>
> As such, [perceptual] experiences are in need of complexity-reducing explanations.
>
> A natural-selection explanation of their occurrence meets the requirement of complexity reduction, and it is not clear that there is any other that does.
>
> The natural selection explanation makes [that is, predicts that—NT] the contents of the experiences it explains [will be] by and large correct [that is, veridical—NT] (98).

The argument, says Peacocke, "does not have the truth of the wholly empirical biological theory of evolution by natural selection as one of its premises." And he supposes that the argument furnishes an a priori case for the claim that our perceptual experiences are by and large veridical.

But consider this analogous argument:

> Planetary orbits are complex events.
>
> As such, planetary orbits are in need of complexity-reducing explanations.
>
> A gravitational explanation of their occurrence meets the requirement of complexity reduction, and it is not clear that there is any other that does.
>
> The gravitational explanation predicts that the planetary orbits it explains will be by and large elliptical.

This argument does not have the truth of the wholly empirical theory of force laws and gravitation as one of its premises. So, if Peacocke were right, we would have here an a priori case for the empirical explanandum that planetary orbits are by and large elliptical! But in the case of perceptual experience we cannot, in the face of skepticism, take as given or empirically confirmed the corresponding explanandum that our perceptual experiences are veridical. For that is precisely what is at issue! Nor can Peacocke claim, in response, that all the premises of his explanatory argument are a priori.

For, first, it does not appear to be a priori that perceptual experiences are complex events. Second, to compound Peacocke's problems, note that 'explain' is a success-verb, and 'explanation', correlatively, a success-noun. Hence the reference, in the conclusion of the argument, to 'the natural-selection explanation' is to an explanation with purportedly true premises—amongst which will be the very empirical claims of evolutionary theory that Peacocke mistakenly believes he has eschewed. He therefore errs in claiming that "[w]hat has been important for the argument is not the empirical truth of Darwinian hypotheses but the special, complexity-reducing status of explanations by some natural-selection mechanism" (108).

Peacocke's invocation of the fundamental insight of evolutionary epistemology is about a quarter of a century late even within the community of Anglo-American philosophers. It goes back to Georg Simmel, and has since been developed as a self-consciously scientific response to philosophical skepticism by writers in the Austro-German tradition, such as Konrad Lorenz, Gerhard Vollmer, and Rupert Riedl.[4]

Perhaps the most important oversight is Peacocke's failure to illustrate or analyze or explicate what he means by a *way of coming to know* a proposition, and, especially, an *a priori way*. His crucial notion of a *contentually a priori content* rests, ultimately, on the notion of an *a priori way*. Yet the latter is taken as an unproblematic primitive, and he proceeds as if it would be clear and acceptable to his reader.

Peacocke offers no epistemology for mathematics. This is puzzling, given the importance accorded to the ontology and epistemology of mathematics by the rationalist precursors whom he cites, and in whose shadows he takes himself to be generating new light. It is not enough to distance oneself from Gödel's "abstract perceptualist" epistemology of number theory and set theory, and allude only to open-textured concepts of abstract objects, and our "implicit conceptions" of the same (180–81). The Gödelian critic has an immediate *tu quoque*. He need only quote Peacocke himself, who does not object to

> cases in which we are, intuitively, inclined to say that it is because *we see the nature* of some kind of entity—a set, a colour, a number, a shape—that we appreciate a priori that certain principles about that entity are correct (181, emphasis added).

---

[4] For further details, see Florian von Schilcher and Neil Tennant, *Philosophy, Evolution, and Human Nature* (Boston: Routledge, 1984).

Peacocke's "resourceful metasemantic theorist" responds to this quasi-perceptual description of the philosophical data by agreeing

> that there is here a special class of examples of the a priori, but he should say that what distinguishes them is as follows. The conditions which individuate the entity in question (the set, colour, number, shape) actually enter the possession-condition for certain canonical concepts of these entities.

This is disappointingly inadequate. What, for example, is the 'canonical concept' of *set*? How does that canonical concept throw any light on the actual structure of the cumulative hierarchy of sets? How does it help to validate the axioms of Zermelo-Fraenkel set theory and its various extensions? What, indeed, are the Peacockean possession-conditions for the concept of *set*? No answers are forthcoming.

One has to provide much more than Peacocke does, in order to be considered a contributor to the great rationalist tradition that derives the greater part of its force from fascination with problems concerning mathematical knowledge. It is also irksome to see the adjectival phrases 'neo-Gödelian' and 'quasi-causal' in this connection. Gödel himself never invoked anything like a causal relation when accounting for our knowledge of mathematical objects. For 'neo' here, one should substitute 'non'.

Peacocke is proficient in neither elementary plane geometry nor the metamathematics of arithmetic. On p. 60, he takes as a priori sufficient for a rectangle to fall under the concept *square* that it be symmetric about the bisectors of its opposite sides. But any rectangle satisfies that condition! What is needed, rather, and will suffice, is symmetry about a diagonal (and indeed just one will do).

There are metamathematical mistakes, of greater moment. On p. 19, Peacocke had advanced a mistaken objection against Conservative Extension of a set of logical rules being necessary for them to determine the meaning of the logical operator involved. His imagined case in point was that adding second-order quantifiers to the language of first-order arithmetic enables one to prove the Gödel sentence of one's first-order theory. He fails to realize, however, that this would not be a case of Conservative Extension. For, upon adding second-order quantifiers, one also helps oneself to all second-order instances of mathematical induction! This is a dramatic extension of the underlying theory. It cannot count as allowing one to prove the new result (that is, the Gödel sentence) only by adding to one's language the logical rules for the second-order quantifiers.

On p. 180, we are invited to "[c]onsider an ordinary person's possession of the concept of a whole number."

> I would say that underlying this person's grasp of the concept is possession of an implicit conception with the content:
>
> (1) 0 is a whole number;
>
> (2) the successor of a whole number is a whole number;
>
> (3) only what is determined to be a whole number on the basis of the preceding two conditions is a whole number.
>
> Now consider the principle that any whole number has only finitely many predecessors. This principle cannot follow from what the ordinary thinker explicitly accepts.

With what justification can one thus thrust on the ordinary thinker only a first-order understanding of (3)? On its obvious reading, (3) is a second-order principle. As such, it guarantees that any whole number has only finitely many predecessors.

The book contains an important self-contradiction. "Contemporary theorists of the a priori should not be involved with the uninstantiated and uninstantiable notion of 'true purely in virtue of meaning' " (27). But "[i]f some principle has an a priori status, its status as such must be explicable in terms of the meaning[s] of the expressions occurring in the principle" (158). All we need here is the very plausible extra premise that some a priori principles are true, and Peacocke is caught in a contradiction. Not too blatant, to be sure, given that the other two premises are separated by 131 pages—but a contradiction nonetheless.

NEIL TENNANT

The Ohio State University

philosophy

NEW FROM CHICAGO

MEMORY, HISTORY, FORGETTING

Paul Ricoeur

Translated by Kathleen Blamey and David Pellauer

A landmark work in philosophy, *Memory, History, Forgetting* examines the reciprocal relationship between remembering and forgetting, showing how it affects both the perception of historical experience and the production of historical narrative. A momentous achievement in Ricoeur's career, this book provides the crucial link between his *Time and Narrative* and *Oneself as Another* and his recent reflections on ethics and the problems of responsibility and representation.

CLOTH $40.00

STATE OF EXCEPTION

Giorgio Agamben

Translated by Kevin Attell

"When a conservative member of the U.S. Congress recently designated the Guantanamo prisoners as 'those who were missed by the bombs' and thus forfeited their right to live, he almost literally evoked Agamben's notion of *homo sacer*, a man reduced to bare life no longer covered by any legal or civil rights. What you hold in your hands is simply *the* book for all those who do not see in 9/11 a mere pretext for patriotic mobilization, but an impetus for a deeper reflection on where we stand today with regard to the very fundamentals of our civilization."—Slavoj Žižek

PAPER $12.00

SARTRE, FOUCAULT, AND HISTORICAL REASON, VOLUME TWO

A Poststructuralist Mapping of History

Thomas R. Flynn

This long-awaited second volume of Thomas Flynn's study of Sartre and Foucault offers a comprehensive and critical reading of Foucault's theories of history.

PAPER $25.00

Also available

SARTRE, FOUCAULT, AND HISTORICAL REASON, VOLUME ONE

Toward an Existentialist Theory of History

Thomas R. Flynn

PAPER $20.00

A DEMOCRACY OF DISTINCTION

Aristotle and the Work of Politics

Jill Frank

"Jill Frank elucidates the substantive arguments in what have looked like paradoxes in Aristotle's thought. She excels at connecting Aristotle's political arguments to current discussions in political theory. Best of all, she elegantly moves beyond the liberal-communitarian debate with an Aristotelian account of politics that not only recognizes the interdependence of individual identity and civic institutions but also respects the human activity of self-creation."
—Danielle Allen, University of Chicago

PAPER $19.00

THE UNIVERSITY OF CHICAGO PRESS
1427 East 60th Street, Chicago, IL 60637 • www.press.uchicago.edu

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2005

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org**

the review of metaphysics

a philosophical quarterly

ISSN 0034-6632
MARCH 2005 | VOL. LVIII, No. 3 | ISSUE No. 231 | $15.00

articles
book reviews
philosophical abstracts
announcements

FREDERICK FERRÉ and respondents	*The Practicality of Metaphysics*
KENNETH SCHMITZ	*Transcendentalism or Transcendentals? A Critical Reflection on the Transcendental Turn*
H. DARREN HIBBS	*Who's an "Idealist"?*
EVGENIA CHERKASOVA	*On the Boundary of Intelligibility: Kant's Conception of Radical Evil and the Limits of Ethical Discourse*
LAURENCE LAMPERT	*Nietzsche's Challenge to Philosophy in the Thought of Leo Strauss*
JUHA HIMANKA	*Husserl's Argumentation for the Pre-Copernican View of the Earth*
ARAN CANES AND STAFF	*Summaries and Comments*

CAMBRIDGE

NEW AND NOTEWORTHY

The Changing Role of the Embryo in Evolutionary Thought

Roots of Evo-Devo

Ron Amundson

Cambridge Studies in Philosophy and Biology

$65.00*: Hardback:
0-521-80699-2: c.280pp.

Autonomy and the Challenges to Liberalism

New Essays

John Christman and Joel Anderson, Editors

$75 00*: Hardback·
0-521-83951-3: c.368pp

Foucault and Classical Antiquity

Power, Ethics and Knowledge

Wolfgang Detel

Modern European Philosophy

$75 00*: Hardback:
0-521-83381-7. c.260pp.

John Locke's Politics of Moral Consensus

Greg Forster

$65.00*: Hardback:
0-521-84218-2: c.328pp.

The Torture Papers: The Road to Abu Ghraib

Karen J. Greenberg and Joshua L. Dratel, Editors

$50.00*. Hardback.
0-521-85324-9: 1284pp.

Kant and the Ethics of Humility

A Story of Dependence, Corruption and Virtue

Jeanine Grenberg

$75.00*: Hardback:
0-521-84681-1. c 260pp.

Naturalistic Hermeneutics

C. Mantzavinos

$55 00*. Hardback:
0-521-84812-1: c.184pp.

Reconceiving Pregnancy and Childcare

Ethics, Experience, and Reproductive Labor

Amy Mullin

Cambridge Studies in Philosophy and Public Policy

$65.00*· Hardback:
0-521-84438-X· c.224pp.

$24.99: Paperback:
0-521-60586-5

Natural Rights Liberalism from Locke to Nozick

Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul, Editors

Social Philosophy and Policy 22

$27.99*: Hardback:
0-521-61514-3: c.424pp.

Wittgenstein's Philosophical Investigations

An Introduction

David G. Stern

Cambridge Introductions to Key Philosophical Texts

$65.00*: Hardback:
0-521-81442-1: c.332pp.

$23 99· Paperback·
0-521-89132-9

Kant's Transcendental Proof of Realism

Kenneth Westphal

$65.00*: Hardback:
0-521-83373-6: c.320pp.

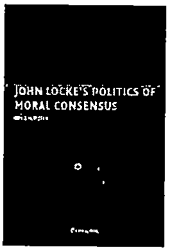
The Selfish Meme

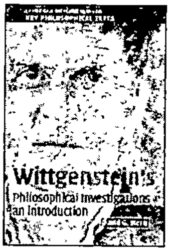
A Critical Reassessment

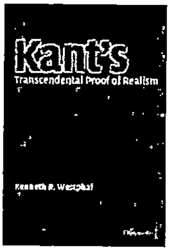
Kate Distin

$55.00*· Hardback· 0-521-84452-5: c.240pp.

$21 99: Paperback: 0-521-60627-6







www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

FROM CAMBRIDGE

The Cambridge Companion to Arabic Philosophy

Peter Adamson and Richard C. Taylor, Editors

Cambridge Companions to Philosophy

$70.00*: Hardback:
0-521-81743-9: c 460pp.
$29.99. Paperback
0-521-52069-X

Is There a Right of Freedom of Expression?

Larry Alexander

Cambridge Studies in Philosophy and Law

$65.00*: Hardback:
0-521-82293-9: c.216pp
$23.99 Paperback:
0-521-52984-0

The Cambridge Companion to Anselm

Brian Davies and Brian Leftow, Editors

Cambridge Companions to Philosophy

$65.00*. Hardback.
0-521-80746-8: c.399pp
$29.99: Paperback:
0-521-00205-2

Now in Paperback...

Lack of Character

Personality and Moral Behavior

John M. Doris

$25.99*: Paperback:
0-521-60890-2: c.288pp

Newton: Philosophical Writings

Andrew Janiak, Editor

Cambridge Texts in the History of Philosophy

$60 00*: Hardback.
0-521-83122-9: c 192pp.
$23.99: Paperback.
0-521-53848-3

Medieval Islamic Philosophical Writings

Muhammad Ali Khalidi, Editor

Cambridge Texts in the History of Philosophy

$65.00*: Hardback:
0-521-82243-2: c 254pp.
$25.99: Paperback:
0-521-52963-8

The Cambridge Companion to Chomsky

James McGilvray, Editor

$75.00*: Hardback:
0-521-78013-6 c.350pp.
$34.99. Paperback:
0-521-78431-X

Now in Paperback...

Faking It

William Ian Miller

$18 99*: Paperback:
0-521-61370-1 c.304pp.

Lessing: Philosophical and Theological Writings

H.B. Nisbet, Editor

Cambridge Texts in the History of Philosophy

$70.00*: Hardback:
0-521-83120-2: c.280pp.
$27.99:Paperback
0-521-53847-5

How the Cold War Transformed Philosophy of Science

To the Icy Slopes of Logic

George A. Reisch

$70 00*: Hardback:
0-521-83797-9: c.430pp.
$26.99: Paperback:
0-521-54689-3

Evidence and Faith

Philosophy and Religion since the Seventeenth Century

Charles Taliaferro

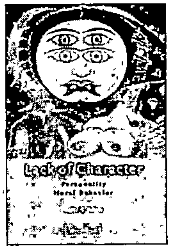
The Evolution of Modern Philosophy

$65 00*: Hardback:
0-521-79027-1: c.330pp.
$29 99: Paperback:
0-521-79375-0

Value and Virtue in a Godless Universe

Erik J. Wielenberg

$55 00 Hardback.
0-521-84565-3: c 200pp. *
$20.99. Paperback:
0-521-60784-1

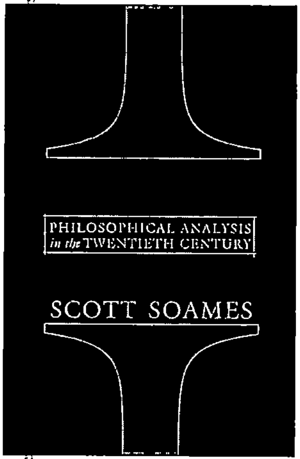


**prices subject to change*

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

Illuminating



New in paperback

PHILOSOPHICAL ANALYSIS IN THE TWENTIETH CENTURY

Scott Soames

"I know of no sustained philosophical work that is as clear, deep, and incisive as these two volumes.... These are superb volumes by a superb philosopher."
—A. P. Martinich, *Journal of the History of Philosophy*

"A history of analytic philosophy and an excellent piece of analytic philosophy in its own right."—Kit Fine, New York University

Winner of the 2003 Award for Best Professional/Scholarly Book in Philosophy, Association of American Publishers

Volume 1: *The Dawn of Analysis*	Volume 2: *The Age of Meaning*
Paper $24.95 ISBN 0-691-12244-X	Paper $24.95 ISBN 0-691-12312-8

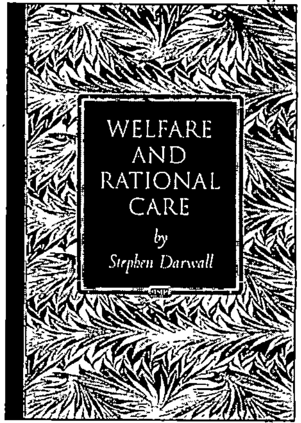
New in paperback

WELFARE AND RATIONAL CARE

Stephen Darwall

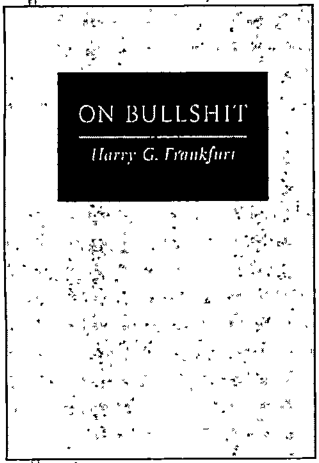
"This book ... covers a great deal of ground in a short space but also has a clear core: an analysis of the key ethical concept of welfare that is both different from and more promising than those already proposed. This is the most interesting analysis of welfare on offer."
—Thomas Hurka, author of *Virtue, Vice, and Value and Perfectionism*

Princeton Monographs in Philosophy
Paper $16.95 ISBN 0-691-09253-2



100 *Celebrating 100 Years of Excellence*
PRINCETON *University Press*
800-777-4726
Read excerpts online
www.pup.princeton.edu

Original



ON BULLSHIT

Harry G. Frankfurt

ON BULLSHIT

Harry G. Frankfurt

One of the most salient features of our culture is that there is so much bullshit. Everyone knows this. But we have no clear understanding of what bullshit is, why there is so much of it, or what functions it serves. In other words, as Harry Frankfurt writes, "we have no theory." With insight and wry humor, Frankfurt attempts to build such a theory.

"A gem of psychological insight, social commentary, philosophical analysis, and good humor.... It is one of the most enjoyable and humanly illuminating short pieces of philosophy produced in the past fifty years."—Raymond Geuss, University of Cambridge

Paper $9.95 ISBN 0-691-12294-6

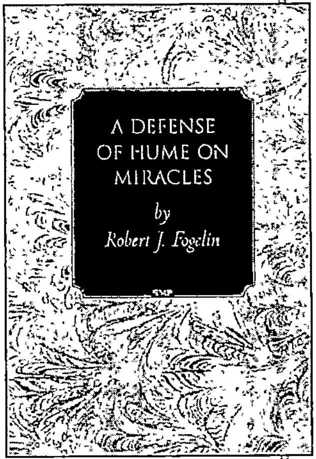
New in paperback

A DEFENSE OF HUME ON MIRACLES

Robert J. Fogelin

"What a joy to read a philosophy book that is graceful, clear, and short.... Fogelin writes with the simplicity and immediacy of a distinguished mind.... [I]mpressively conceived and executed." —Mark Sainsbury, *Times Literary Supplement*

Princeton Monographs in Philosophy
Harry Frankfurt, series editor
Paper $19.95 ISBN 0-691-12243-1



A DEFENSE
OF HUME ON
MIRACLES
by
Robert J. Fogelin

New from
Oxford

Truth, Thought, Reason
Essays on Frege
TYLER BURGE
Tyler Burge presents a collection of his seminal essays on Gottlob Frege (1848–1925), who has a strong claim to be seen as the founder of modern analytic philosophy, and whose work remains at the center of philosophical debate today. *Truth, Thought, Reason* gathers some of Burge's most influential work from the last twenty-five years, and also features important new material, including a substantial introduction and postscripts to four of the ten papers.
March 2005 paper $24 95 cloth $90.00

How the Body Shapes the Mind
SHAUN GALLAGHER
Shaun Gallagher uses the concepts of body image and body schema to address a range of philosophical questions about cognition, drawing on recent work in developmental psychology, brain research, and studies of unusual pathologies. Truly interdisciplinary, the book will reward academics and professionals in philosophy, psychology, neuroscience, and beyond.
March 2005 $55 00

New in Paperback!
Aspects of Reason
the late PAUL GRICE
introduction by RICHARD WARNER
Here is the long-awaited publication of the John Locke Lectures on reasons and reasoning that Grice delivered in Oxford at the end of the 1970s. The lectures focus on an investigation of practical necessity, placing the concept in the context of an account of reasons and reasoning.
paper March 2005 $27.50

New in Paperback!
The Practice of Value
JOSEPH RAZ
Edited by R. JAY WALLACE,
Commentaries by CHRISTINE M. KORSGAARD,
ROBERT PIPPIN, AND BERNARD WILLIAMS
The Practice of Value is an exploration of a pervasive but puzzling aspect of our world: value. The starting-point is the Berkeley Tanner Lectures delivered in 2001 by the leading moral theorist Joseph Raz. His aim is to make sense of the dependence of value on social practice, without falling back on cultural relativism.
paper March 2005 $16 95

Reality and Rationality
the late WESLEY C. SALMON
Edited by MERRILEE H. SALMON and PHIL DOWE
This volume of articles is a follow-up to the late Wesley C. Salmon's widely read collection *Causality And Explanation* (OUP 1998). It contains both published and unpublished articles, and focuses on two related areas of inquiry: First, is science a rational enterprise? Secondly, does science yield objective information about our world, even the aspects that we cannot observe directly? Salmon's own take is that objective knowledge of the world is possible, and his work in these articles centers around proving that this can be so.
March 2005 paper $29.95 cloth $74 00
JKC 121004

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us.

OXFORD
UNIVERSITY PRESS

Pune University
patta
7·11·05
[illegible]
CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 4
APRIL 2005

page
163 *Wisdom and Perspective* Valerie Tiberius
183 *The Dif* Kadri Vihvelin and Terrance Tomkow

BOOK REVIEWS
206 Philip Kitcher: *In Mendel's Mirror: Philosophical Reflections on Biology* Mohan Matthen

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022-362X

CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NO. 4, APRIL 2005

WISDOM AND PERSPECTIVE*

In Ann Patchett's novel *Bel Canto*,[1] the Vice President of a small Latin American country and about fifty others are held hostage in the Vice President's mansion. They have been cooped up in the living room for months and many believe they will not survive the ordeal. When their captors finally let them outside for some air:

> Vice President Ruben Iglesias, who thought he would not live to feel once again the sensation of grass beneath his feet, stepped off the shale stone walkway and sank into the luxury of his own yard. He had stared at it every day from the living-room window but now that he was actually there it seemed like a new world. Had he ever walked around his own lawn in the evening? Had he made a mental note of the trees, the miraculous flowering bushes that grew up around the wall? What were they called? He dropped his face into the nest of deep purple blossoms and inhaled. Dear God, if he were to get out of this alive he would be attentive to his plants (*ibid.*, p. 281).

Ruben Iglesias undergoes a change in perspective on his life as a result of the highly unusual circumstances in which he finds himself. Instead of valuing success in his career and working for the sake of that goal at the cost of many other possible pursuits, as he had done for many years, he begins to see the importance of stopping to smell

---

* I would like to thank Bridget Clarke, Roger Crisp, Dan Haybron, Rosalind Hursthouse, Mark LeBar, Jimmy Lenman, Elijah Millgram, George Sher, Karen Stohr, Christine Swanton, J.D. Walker, R. Jay Wallace, Gary Watson, Jennifer Whiting, the Department of Philosophy at the University of Sheffield, the audience at the Rocky Mountain Virtue Ethics Summit in Boulder, Colorado, 2004, and the Ockham Society at Oxford University, for helpful discussion and comments on earlier versions of this paper. I would also like to thank the Graduate School at the University of Minnesota for the McKnight Land-Grant Professorship, which provided support for the research on this paper.

[1] New York: HarperCollins, 2001.

0022-362X/05/0204/163–82

the roses, in his case quite literally. Several of the other characters in the novel experience similar changes in perspective, as they come to appreciate the value of things they previously took for granted or did not have time to notice at all.

Though few of us have been in such a threatening situation, I suspect that most of us would recognize the experience of a perspective shift. Major life events trigger shifts in perspective and minor shifts in our attention to values happen frequently. In this paper I argue that shifts in evaluative perspective are vital for living a prudentially good life and that this fact has important implications for how we conceive of practical wisdom. One way to accommodate shifts in evaluative perspective would be to adopt an extreme form of pluralism according to which we pursue incommensurable values without any overarching plan or conception of a good life that brings these values into order. But dismissing the conception of a good life that is part of traditional philosophical accounts of living well is a drastic move and one that we have reason to avoid. I argue instead that consideration of perspective shifts ought to change the way we understand the notion of a conception of a good life, but that this notion should not be entirely abandoned. Since one vital role of practical wisdom has to do with the construction and application of our conception of a good life, this argument will enrich our understanding of practical wisdom as well.

I. THE VALUE OF SHIFTING PERSPECTIVES

I suspect that many of us, reflecting on Ruben Iglesias's change of perspective, will think that this is a valuable experience, one that it would be good for many of us to have if we could have it without the risk of death. Acknowledging the beauty of our everyday surroundings and appreciating the wonders of nature are good things to do and sometimes an invaluable tonic for the materialistic or accomplishment-focused perspectives that can easily absorb us. Nevertheless, it would not obviously be a good thing for Ruben to keep his attention focused on his plants, to the exclusion of other interests and concerns, especially once he is released from being a hostage and his duties as Vice President resume. It is good to stop and smell the roses, but not so good to smell them all the time.

Ruben Iglesias is not the only example we can find to illustrate the value of shifting perspectives on life. Nor is it necessary to have one's life under the imminent threat of death in order to experience such shifts. There are many events that can cause us to see life in a different way, to change our priorities or our values, even if only temporarily. The death of a loved one can have this effect, as can other major life changes such as being fired from a job, surviving a major illness, or having a child. Changing one's physical surroundings by traveling or getting out into nature can trigger changes in perspective, as can an encounter with great art. There are also ways one can bring about such a change intentionally, for instance by meditating or just going for a walk. Further, the two perspectives that Ruben Iglesias's case draws to our attention—one focused on career success and the other on natural beauty—are not the only two perspectives one can have. The unexpected death of an acquaintance can make us take the perspective of "living for the moment" in which short-term pleasures seem paramount, whereas the arrival of a child can make us take a perspective that emphasizes the long-term benefits of our actions to a wider circle of people.

None of the perspectives just mentioned is a paradigm of the rational or reflective perspective recommended for making normative judgments. Ruben Iglesias is not focused on evaluating his reasons for valuing nature in light of the facts. Rather, he is in the grip of a set of attitudes and dispositions to act. He is, I will say, absorbed in a particular *practical perspective*. From a particular practical perspective one set of values plays the role of a goal for action and appropriate feeling, and critical scrutiny of these values is suspended. From a *reflective point of view*, on the other hand, the point is to engage in critical scrutiny about some of our values. The result of taking a reflective point of view is not (at least not immediately) action or emotion, but considered judgments about our conception of a good life. To clarify, reflection here does not mean intellectual engagement in general. Being intellectually engaged, say, by a puzzle or a philosophical problem, is being in a practical perspective that highlights the value of truth, intellectual achievement or the like. The kind of reflection I mean to exclude from occupying practical perspectives is the much more specific kind of critical reflection that concerns the justification of one's values or projects.

In this section I want to make the case for the claim that shifts from one practical perspective to another, and between practical perspectives and reflection, are a valuable and important part of a good life. As we shall see, these shifts are important in two related ways: First, shifting perspectives is necessary for the full realization or pursuit of the values they highlight. Second, this realization of values informs our reflection about the shape our conception of a good life ought to have.

The first point is nicely illustrated by the case of Juan and Linda from Peter Railton's well-known paper.[2] Juan and Linda are a happily committed couple who live in two different states. Linda is depressed and Juan knows that an extra visit from him would help. So he goes, even though the money he spends on this visit could be much more helpful to others if he were to give it to Oxfam. Because Juan is a "sophisticated consequentialist," Railton argues, he will not try to perform the most beneficial action, rather, he will act on the disposition he has cultivated for consequentialist reasons, namely his love for Linda. According to Railton, "in thought and action we shuttle back and forth from more personal to less personal standpoints and both have an important role in the process whereby identity, meaning and purpose are generated and sustained" (*ibid.*, pp. 164–65). Juan's moral point of view is consequentialist: he thinks that he ought to do whatever will produce the best consequences for all concerned. But Juan also thinks that a world without loving relationships would be unbearable and so he fosters dispositions in himself that allow for real loving relationships even though they may cause him to violate his own criterion of right action. In order to have a truly loving relationship (to meet that goal of his) Juan must be able to shift to a perspective from which Linda's happiness is most important.

Railton's move to sophisticated consequentialism is supposed to explain the way in which consequentialism is less alienating than critics have charged.[3] His own view is that this move does not result in multiplying perspectives or points of view for the subject. Rather, he seems to think that there is one unified point of view that comprehends both the value of maximizing objective goods and the value of Linda's happiness. But I think this is only partly true. While Railton may be right that the philosophical perspective from which Juan grasps the consequentialist criterion of right action and the justification for his partial dispositions toward Linda, the perspective from which he decides whether to take the flight is not all encompassing in this way. When he decides to take the flight his love for Linda takes over and he does not, as Railton says, "even try to do the most beneficial thing" (*op. cit.*, p. 159). A comprehensive reflective point of view may be available to Juan, but he cannot stay in this perspective and hope to achieve the values that it recommends. When he makes decisions about how to treat Linda he cannot *at the same time* and in every case be wondering whether his life is morally defensible. If he does this he has retreated to the kind of consequentialism that is truly alienating and he has missed out on an important element of a good human life.

---

[2] "Alienation, Consequentialism and the Demands of Morality," *Philosophy and Public Affairs*, XIII (1984): 134–71.

[3] For such criticism see Bernard Williams, "A Critique of Utilitarianism," in Williams and J.J.C. Smart, *Utilitarianism: For and Against* (New York: Cambridge, 1973), pp. 108–18; and Michael Stocker, "The Schizophrenia of Modern Ethical Theories," this JOURNAL, LXXIII, 14 (August 12, 1976): 453–66.

One might think that it is only if the reflective point of view in question is consequentialist that one needs to become absorbed in particular practical perspectives.[4] But this is not the case. To see that the point has broader scope, consider the case of someone who takes up a reflective perspective that is, more or less, virtue ethical, centered around the nature of a flourishing life for a person. Let us imagine that Ruben Iglesias is such a person. From a reflective point of view, Ruben has a comprehensive conception of a good human life that includes a variety of worthwhile goals, each of which emphasizes a different aspect of his nature. This conception of the good life includes judgments about the reasons for developing particular virtues and pursuing particular ends such as friendship, health, and mental cultivation. From the reflective point of view, Ruben can see that the justification for pursuing the particular ingredients of his conception of the good life has to do with what it is for him to flourish as a human being.

Now if Ruben is going to have the kinds of attachments to his wife and children that are necessary for his flourishing, his attachment to them needs to transcend this perspective. He needs to be devoted to them in a way that has nothing to do with his flourishing. Of course virtue ethics does not claim that people should value friendship and other important ends *for the sake of* their own flourishing; on the contrary, part of what it is to have the virtues relevant to friendship, for example, is to love friends for their own sakes. But virtue ethics also typically maintains that developing the virtues is part of what it is for an individual to flourish. So, far from eliminating the divide between perspectives, virtue ethics appropriately highlights the existence and importance of multiple perspectives. The perspective from which Ruben can grasp a comprehensive conception of the good is not the perspective from which he can realize the particular values that comprise this conception. To be an effective and dedicated politician, he has to be devoted to his job so that at certain times it seems the most important thing in the world to him. Similarly, to be a good father, he must respond lovingly, without stopping to reflect or think about the reasons for which being a good father is valuable. Moreover, the pattern of dispositions to action and emotion that make Ruben thrilled to be in politics is not likely to be the same one that makes him a devoted father. To gain all that he can out of these parts of his life, he needs to shift between different practical perspectives.

---

[4] An implication of the discussion here is that a person may have more than one reflective perspective. A person might have a consequentialist moral perspective that is reflective in the sense that it requires a stepping back from particular commitments in order to think about the justification of one's moral commitments, and a reflective perspective on what it is for her own life to go well that is not particularly consequentialist. On my view, the reflective perspective on how to live encompasses the moral because moral principles and values are one important set of commitments that matters to how well our lives go. The claim is not crucial, however, for the point that multiple practical and reflective perspectives are necessary for living well.

The case of virtue ethics makes it easy to see how even deep reflection on the good life for us is not a point of view we can occupy all the time. In short, taking a reflective point of view on our values is not the same as being engaged by them and we cannot be fully engaged by everything that has value for us at once. Thinking abstractly about what has value and what ends are important to human life, we may very well be able to acknowledge intellectually all the various important ends and values. My point is not that there are truths about value that are impossible to articulate or grasp in any way from the reflective point of view.[5] But engaging in the pursuit of ends or obtaining an emotional grasp of certain values is not an abstract intellectual exercise. It is here in the everyday business of living that we need shifts in perspective.

One conclusion we might draw from the distinction between reflective and practical perspectives is that the practical perspectives are an unfortunate but necessary compromise that psychologically limited beings like us must make. The reflective point of view is the one from which we see the truth about what has value, but we can only pursue these values if we forget this reflective point of view and allow those values to reign over us temporarily. But this conclusion ignores what I take to be a very important fact about practical perspectives, namely, that it is from within these perspectives that we learn about what has value and that some of what there is to learn can only be learned by being absorbed by the values in question. This brings us to the second claim I wanted to make about shifting perspectives, which is that practical perspectives inform the reflective perspective in an important way. Because being absorbed in a practical perspective with our undivided attention is sometimes necessary for achieving the values inherent in it, practical perspectives are a source of information about the valuable goals of human life.

---

[5] I am grateful to Jimmy Lenman for helpful discussion on this point.

The informative function of shifts in perspective is made evident by the phenomenon of being in a rut. It seems that we can get stuck in a practical perspective in such a way that we are prevented from achieving a good life. When this happens, we need to discover or be reminded of other important values and goals. Sometimes a reflective point of view from which we see the real value of each thing and do not get carried away by any one thing is helpful. (This reflective point of view helps us put things in the right perspective by revealing that what we are distressed about is not important enough to warrant our reaction.) But sometimes a completely different practical perspective that immerses us in other values is more valuable than measured reflection.[6] Practical perspectives inform our reflection because it is (at least in part) by being a friend, daughter, sibling, or parent that we discover what is valuable about these relationships. It is by absorbing ourselves in a hobby or career that we experience the value of accomplishment. It is by losing ourselves in the moment that we experience the value of pleasure, peace of mind, or fun.

Another kind of example will help to illustrate the importance of this kind of absorption to the informative function of practical perspectives. Some activities such as rock-climbing, playing the violin, or solving a philosophical problem, absorb us and demand all of our attention. Would the rock-climber be better off if she were not so limited that climbing rocks required all her attention? Would she be better off if she could climb rocks while composing a shopping list, talking to her children, planning her garden, and reflecting on the meaning of life? There may be times in life when such multitasking is desirable, but think of what is missed if we are always multitasking in this way. "Being in the moment," although now a cliché, does have much to recommend it. Experiences such as the awe of nature, the physical exhilaration of dance, sex, or sport, or the mental exhilaration of a great conversation are not experiences we can really have while wondering what the value of them is and how they fit into our lives.

Moreover, examples such as these reveal that it would not be desirable, even if it were possible, to occupy many practical perspectives at once. While it is true that some values can be pursued or appreciated together to mutual advantage, not all practical perspectives are mutually supporting in this way. Trying to enjoy sex while appreciating the beauty of nature may very well frustrate both aims and a great conversation can distract one from an appreciation of art if the conversation is about something else (or sometimes even if it is not).

---

[6] Another character in Patchett's novel, a reserved and responsible businessman, discovers the joys of romantic love when threatened with the possibility of imminent death. This new perspective could not have resulted from reflection because he did not know that romantic love was something he was missing; the new perspective teaches him something vital about life that he could not have discovered by thinking about it.

To sum up, we cannot at the same time occupy the reflective point of view and be fully absorbed in a practical perspective, nor can we occupy many practical perspectives at once. If a good life includes the realization and appreciation of many different values in addition to deep reflection, then shifting perspectives is necessary for a good life. Eliminating these shifts would leave us with vastly diminished resources for learning about what has value and for achieving the ends that we know are valuable. If this is right, then it seems that we need to be able to negotiate changes in perspective and this leads us to the domain of practical wisdom.

II. LIVING WELL AND THE CONCEPTION OF A GOOD LIFE

Before I say more about the implications of my discussion for an account of practical wisdom, it is worth considering an alternative response to these observations about shifting perspectives. In the above discussion I have assumed that appropriate reflection is a part of a prudentially good life and this assumption might now seem dubious. Why not just abandon the kind of reflection that interferes with our genuine immersion in values? Though too much reflection or reflection at the wrong time frustrates our living well, not to reflect at all would also be a mistake. When we reflect on what it is for our lives to go well we form a conception of a good life. Because our various commitments provide support for each other, a conception of a good life that locates these commitments on the same map ensures that the commitments we have do not undermine each other and can be pursued together in the same life. A reflective conception of a good life, then, situates our individual commitments in a justificatory framework. This framework serves at least two important purposes. First, it makes it more likely that the values we have can be pursued together in the same life. Second, it meets our need for our own reflective approval of how our lives are going. Without such a conception we would have no sense of how our various commitments function together as an evaluative standard and no reason for confidence in the justification of the individual commitments we have. A reflective perspective on our own good gives us a reasonable set of standards that make possible the satisfactory review of our own conduct that is part of what it is to live well.[7]

---

[7] The view that a satisfactory self-survey is an essential component of living well is compelling and widely held. Hume suggests that bearing one's own survey as an

A conception of a good life is more than just a jumbled set of value commitments. The conception of a good life imposes some structure and order on these commitments. In particular, a conception of a good life locates the various value commitments one has and reveals how they are related with respect to mutual support and relative priority. Further, our conception of the good life will include a judgment about the relative importance of the plural values that constitute it. This is not to say that our values will be ranked on a cardinal scale, but insofar as some things are clearly more important to us than others this fact will be reflected in our conception of a good life.

A conception of a good life is structured, then, in the sense that it locates the components of a good life with respect to each other in some way, but it need not have a particular structure such as the structure of a life plan, nor should the structure be rigid and immutable.[8] Conceptions of a good life cannot be too rigid or detailed because these conceptions are also informed by experience gained from living life when we are not reflective. As we have seen, nonreflective experience is a crucial component of a good life and a vital source of information about what it is to live well. Our conception of how to live, then, must be flexible and open to change.

A conception of a good life that serves these goals requires reflection about the nature and justification of a conception of a good life. As we have seen, we cannot engage in deep reflection all the time, nor is a good life one in which we are always disposed to engage in such reflection.[9] Given this, we need to know when is it appropriate to be

---

important goal in human life when he embarks on a brief exhortation to virtue at the end of the *Treatise*—see *A Treatise of Human Nature*, L.A. Selby-Bigge, ed. (New York: Oxford, 1978, second edition), pp. 619–20. Other philosophers who share this assumption include Rüdiger Bittner, *What Reason Demands*, Theodore Talbot, trans. (New York: Cambridge, 1989), p. 123; Thomas E. Hill, "Pains and Projects," in his *Autonomy and Self-Respect* (New York: Cambridge, 1991), pp. 173–88; John Rawls, *A Theory of Justice* (Cambridge: Harvard, 1971), p. 422; and Charles Taylor, "Responsibility for Self," in Amelie Rorty, ed., *The Identities of Persons* (Berkeley: California UP, 1976), pp. 281–99.

[8] For an interesting discussion of the idea that the plan model is not the only way of conceptualizing a life, see Michael Walzer's *Thick and Thin: Moral Argument at Home and Abroad* (Notre Dame: University Press, 1994), pp. 23–24. In a later chapter titled "The Divided Self," Walzer distinguishes "divided selves," which he thinks most of us are, from pathological "utterly fragmented selves" (p. 98). As will become more clear in this paper, I am very sympathetic to the idea that we are divided selves. By distinguishing order or structure on the one hand from constancy, inflexibility, or detail on the other my account of the good life is meant to be one that is accessible to divided selves. I first read about this discussion in Walzer in Carl Elliott's *Better Than Well: American Medicine Meets the American Dream* (New York: Norton, 2003), which also contains an interesting discussion of the plan model of a good life (p. 299).

[9] Henceforth, I will use "reflection" to refer to this kind of deep, critical reflection. The claims I make about reflection in my sense should not be taken to be true of a broader conception one might have that would include any kind of cognitive process.

reflective. Here we find that traditional accounts of practical rationality are not very helpful. Philosophers have tended to recommend that in a reflective moment we deliberate, make a life plan, think about our conception of the good, or decide which ends to endorse, and then we put this plan, these decisions, or these choices into action.[10] The direction of rational authority is top down: the plans, choices, and judgments we make when we are reflective determine the rationality of the choices, actions, and feelings we have in practice. Of course, such accounts acknowledge the distinction between reflection and practice. But these accounts of practical rationality aim to characterize the rationality and the reflective point of view by articulating the principles or standards that govern practical reason. They do not take movement between reflection and other perspectives, nor certainly movement between perspectives that are not reflective, to be part of the province of a theory of practical reason.

Aristotelians have been better about not assuming a top-down picture.[11] Nevertheless, even Aristotelian conceptions of wisdom could benefit from an acknowledgment of the role of shifting perspectives in a good life. The ideal of the unity of the virtues in the practically wise implies that everything we need to grasp in order to act well can be appreciated from the reflective point of view and this is an assumption that is challenged by the preceding account of the value of perspective shifts. Sabina Lovibond's characterization of practical wisdom, for example, includes both a grasp and an ordering of all that matter in life.[12] If I am right that we learn about values from our engagement in practical perspectives, then an overarching point of view that specifies the relative value of each constituent of the good life may not be the right goal. While there is a sense in which we can acknowledge a plurality of values from the reflective point of view, this point of view may not be the one from which we have the kind of appreciation of values that allows us always to act in a way that is conducive to a good life.

---

[10] According to one very popular view of practical reasoning, instrumentalism, our ends are given by our desires and reasoning tells us how to pursue them. On this view it is also true that the scope of reasoning is limited to a particular perspective: the perspective of matching means to ends.

[11] Martha Nussbaum, for instance, emphasizes the fact that practical reasoning for Aristotle is not a science and that it is concerned with "insight through experience"—see her *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (New York: Cambridge, 1986, revised edition), p. 299. Sarah Broadie argues against the top-down or, as she calls it, the "Grand End" picture of practical wisdom in her *Ethics with Aristotle* (New York: Oxford, 1991), pp. 198–202. See also John McDowell, "Deliberation and Moral Development," in Stephen Engstrom and Jennifer Whiting, eds., *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty* (New York: Cambridge, 1996), pp. 19–35.

[12] *Ethical Formation* (Cambridge: Harvard, 2002), pp. 27–29. I take Lovibond's account of wisdom to be sympathetic to mine insofar as she emphasizes what she calls "the openness to the layout of reality" (p. 24) or what I would call an openness to evaluative perspectives.

In the remainder of this essay I argue that the top down picture of practical rationality is incomplete and that this matters for how we characterize the virtue of practical wisdom and the reflective agent.[13]

III. WISDOM AND PERSPECTIVE SHIFTS

We need to be able to shift from one practical perspective to another, and to take a reflective point of view on our life as a whole when appropriate. Importantly, we need to be able to make these shifts without the top down application of a comprehensive reflective model. First of all, the choice to shift perspective cannot be the result of applying the reflective model to practice in any rigorous sense because one of the things that needs to be judged is when to occupy that reflective point of view. In other words, we cannot decide when we need to shift perspectives by becoming reflective and judging that our model of a good human life implies that we ought to make a change. This mode of change assumes that we have already decided to become more reflective. The basis for this original shift in perspectives cannot be a direct application of a reflective model.

Second, even when it comes to shifts between practical perspectives, we cannot always decide about these shifts by occupying a reflective perspective and standing in judgment. This is because of the way in which practical perspectives have us in their grip. To be in a particular practical perspective is to have your emotional responses and dispositions to act accord with the values that define that perspective. Practical perspectives also shape the everyday practical reasoning, planning, and decision making that we do on the basis of (temporarily) fixed goals or values. Practical perspectives have a certain life of their own, an inertia that is the result of these emotional and dispositional patterns. When Juan is living his life, being a good husband and not engaging in reflection about the permissibility of his projects, his love for his wife makes certain courses of action seem obvious and it crowds out other options and other ways of responding to her distress. The reflective perspective is not always available to us when we could most benefit from a shift in perspectives. So, we need to be able to judge that we should try to see things differently without already having taken up a reflective stance. One could, of course, stand perpetually ready to engage in deep reflection, always aware of the way in which one's practical commitments are contingent on reflective approval, always ready to evaluate one's commitments from the reflective point of view. But a person who lives this way does not gain what there is to gain from being absorbed in practical perspectives.

---

[13] Practical rationality and practical wisdom are not the same thing, but I take it that the capacity for practical reason is one part of practical wisdom. The claim here that wisdom does not always require being rational or reflective (where this implies a detached perspective) does not mean that the calm, cool, deliberative moment is never the right perspective to have. The capacity to disengage from our passions and deliberate calmly is an important one, but I will argue that it is not the only capacity that comprises practical wisdom.

Reflecting on our examples, it might seem that shifts in perspective are caused by changes in external circumstances. In Ruben Iglesias's case the shift is forced by drastic changes in his circumstances. In many other cases, shifts in perspective seem to be the natural result of ordinary shifting circumstances: the rock-climber does not have to make an effort to focus her attention on the rocks and forget about her job. Similarly, coming home from the office tends to shift a person's attention away from values associated with his job and toward the values associated with family and with being a parent or partner.

This explanation makes it seem that shifting perspectives is not something we do, but something that happens to us. And this may make it seem unlikely that there is anything to say about how we shift perspectives without taking up the reflective point of view and even less likely that shifting perspectives has anything to do with wisdom. It is true that perspective shifts can be caused by external changes, nevertheless, dismissing the role for agency and the virtue of practical wisdom is unwarranted. We can see why if we think about two kinds of failure with respect to perspective shifts. First, there are cases in which people's perspectives do not change, despite the change in external circumstances. Some people who come home from work do not stop thinking about the office. And some people who hike, play music, or sit on the beach watching the sunset never become fully absorbed by the experience. Second, there are cases in which perspective shifts occur inappropriately. Sometimes a shift in perspective is really a way of avoiding or retreating from something important. People sometimes become reflective when they ought to be enjoying the moment due to fear or self-doubt, and others avoid reflection when it would be appropriate due to a desire to avoid a difficult decision.

What takes wisdom is shifting perspectives at the right time, in the right way and for the right reasons. The wise person is open to perspective shifts and the reasons for them and, consequently, shifts perspective when it is appropriate to do so. We can understand what it is to be open to shifts in perspective as a capacity to grasp reasons or values, quasi-intuitively, without engaging in any reflection on how they are justified. This openness to reasons and values is a capacity that can be more or less developed. It can vary in its tendency to grasp the right things and in its tendency to grasp them at the right time. The fact that our power to grasp reasons and values can be better or worse makes room for the virtue of wisdom. The wise person grasps the right reasons and values at the appropriate time and changes her perspective accordingly.

This notion of shifting perspectives for the right reasons creates a problem because the obvious way to understand the notion of *right* here is by appeal to a person's reflective conception of the good life. Now there are some ways in which a person can appeal directly to her conception of a good life in order to effect changes in her perspective without actually having to take up that point of view when a change in perspective is needed.[14] For example, from a reflective point of view, recognizing our tendency to get stuck in a certain practical perspective, we can engage in self-manipulation or pre-commitment to effect changes at a later point in time when we are no longer being reflective. Consider the person who recognizes her tendency to bring her work home with her and plans to have a long bath and a cocktail when she gets home to put her in a different frame of mind. In this case the plans she makes while reflective have some influence on perspective changes later.

Self-manipulation and pre-commitments are sometimes foiled, however, as for example when we cannot see the reason for following the plan we adopted from a reflective point of view once we are out of that point of view. Moreover, these strategies are not available for every kind of desirable shift. Sometimes the conditions that give us a good reason to change our perspective are not predictable and are not anticipated from the reflective point of view. The wise person, therefore, must sometimes make such shifts using only the resources available from within a particular practical perspective, without referring to her reflective conception directly. And this presents a problem given that it is from the reflective point of view that we consider and evaluate reasons.

The answer to the problem consists in two claims: first, other normative considerations are available from within particular practical perspectives. While it is true that a fully reflective perspective intrudes upon an engaged practical perspective, a person within a practical perspective is not entirely blind to considerations that present reasons for shifting out of that perspective. Competing values are less salient but not necessarily entirely absent from within a practical perspective and therefore the wise person can recognize the force of other values from within a practical perspective.

---

[14] I thank Elijah Millgram for alerting me to this possibility.

The second part of the answer is that we can grasp and act on these normative considerations without occupying a reflective point of view. The wise person does make an appraisal of the reasons there are for shifting perspectives, but this appraisal is not made by applying a decision procedure or model to the circumstances. Rather, the wise person has an openness to the intuitions, feelings, and perceptions that draw her attention to the relevant reasons without engaging her capacity for deep critical reflection. Being open to reasons from within a practical perspective means being able to appreciate what is at stake in considerations that are not at the center of attention from that perspective. Since considerations may appear to us as reasons without bringing along the justificatory background that makes them reasons, our acknowledgment of these reasons need not invoke a reflective conception of the good life nor require that we take up a reflective point of view. This appreciation of reasons is not an explicit rational acknowledgment but something more like an intuition or impression.

Such judgments about reasons for shifting perspectives, based as they are on an intuitive grasp of reasons, need not take us to the reflective point of view. In order to see how this is so, let us consider an example of a particular type of judgment, one that has a natural role in shifts of perspective, namely, judgments about the character manifested by occupying a particular perspective. Consider Gus who is very devoted to his dog. Gus believes that his relationship to his dog is valuable and important. Caring for another entirely dependent being has taught him compassion and has allowed him to extend his sympathetic capacities.[15] Observing the dog's way of being in the world has also encouraged him to enjoy life in ways he did not before. While Gus's relationship with his dog is, by and large, healthy, he has a tendency to become obsessed with concerns about his dog's welfare to the point of distraction. During one of these times, Gus worries so much about Max at work that he cannot get anything done and he declines invitations he would like to accept because he is so worried about leaving Max alone. We might say that if Gus were wise, he would recognize that being stuck in this perspective where Max has taken complete priority is obsessive, self-indulgent, or melodramatic. He may think his perspective is obsessive because it is preventing him from achieving many of his goals. He may think it is self-indulgent if he recognizes that he is using it as an excuse to avoid social engagements or challenges at work that he finds intimidating. Or, he might assess his occupying this perspective as melodramatic if he realizes that he is really looking for attention and that his concerns for Max have not translated into any actions that are particularly good from the dog's point of view.

---

[15] Lori Gruen argues that because animals are so different from us, our relationships with them help to develop and extend our capacities for sympathy, empathy, and compassion.

Here it is natural to say that it "registers with" or "dawns on" Gus that he is being obsessive. Or we might say that Gus grasps, quasi-intuitively, or impressionistically that his attention to his dog has become pathetic or unhealthy. In these cases the wise Gus is *open* to the reasons for shifting from his "dog-centric" perspective and they come to him through intuition and feeling rather than explicit thoughts about reasons.[16] His judgment that his perspective is obsessive, melodramatic, or the like is based on his not fully reflective grasp of these reasons.

Similarly, Juan recognizes that letting his personal perspective crowd out the global perspective from which his wife's needs diminish in importance would be uncompassionate or blind. He can see this from the perspective in which his wife's needs matter most and it is what keeps him from becoming obsessively devoted. Losing the personal perspective altogether, though, would be insensitive or unloving and this can be grasped from other points of view. Notice that these virtue evaluations capture the ways in which perspective shifts are valuable and that the same considerations might represent themselves in terms of the values that are at stake. To call something obsessive is, in part, to say that you have focused on it to such an extreme extent that it is preventing you from achieving the other things you value. To say that you are being insensitive, blind, or narrow-minded is, in part, to say that your current perspective prevents you from noticing other ethically salient features.

A wise person shifts perspective when there is reason to do so. Because the resources for shifting perspectives must be ones that are available from within a particular engaged practical perspective, the wise person must have the ability to appreciate reasons for shifting perspectives that do not require reflective engagement or explicit rational thought. The practically wise person, then, is open to normative considerations that are not the focus of her current practical perspective and she is able to make judgments on the basis of these considerations about her current perspective.[17]

---

[16] I am grateful to Jennifer Whiting and George Sher for very helpful discussion here.

IV. WISDOM AND RATIONALITY

The wise person, according to Aristotle, is able "to deliberate well about what is good and expedient...about what sort of things conduce to the good life in general."[18] If shifting between the reflective point of view and different practical perspectives that focus our attention on different values is part of a good life, then wisdom includes the capacity to make such shifts appropriately. I have argued that appropriate shifts do not employ inferential practical reasoning. Given the close relationship between wisdom and practical reason, this may be quite jarring. In this section I take up some objections to the association of wisdom with the quasi-intuitive process I have described.

First, some might be concerned that the admission that judgments about shifting perspectives are not directly inferred from one's conception of a good life introduces too much looseness into the picture of practical wisdom. After all, if the guidance provided by the conception of a good life is indirect, the mapping between that conception and how one actually lives will admit of gaps. Notice, though, that these gaps cannot be eliminated if we are going to learn what practical perspectives have to teach us. There is no way of life at which we can aim that is a perfect mapping of reflection onto practice. Or, at least, such a life would not really be ideal. A neatly mapped life in which we realize our ideal conception of how to live well is a life in which we are always somewhat detached, always ready to check ourselves against an ideal model; it is a life in which we do not get lost in experience and therefore do not learn from it.

A more serious concern has to do with the justification of the intuitive grasp of reasons that provides the basis for judgments about the need to shift perspective. That is, one might be concerned that if these judgments about one's practical perspective are not made in accordance with a rational process, if they are not directly inferred from one's conception of a good life, then they will not be justified at all. But this concern ignores the justificatory force that such impressionistic judgments do have. Our ordinary normative judgments, made without full reflection, have some authority in virtue of the fact that our reflection on the good life must be informed by experience.[19] The authority of these judgments increases as they become integrated into a more reflective point of view and become part of a refined set of judgments, but the process of justifying our conception of a good life and our particular normative judgments is a dialectic process that takes seriously the verdicts of critical reflection and the teachings of experience.

---

[17] As we have seen, a wise person also has a conception of her own good. Even with this addition, the account here is not a complete account of practical wisdom. A complete account would also seem to include, at least, moral knowledge, which I have not discussed.

[18] Aristotle, *Nichomachean Ethics*, translated by W.D. Ross, revised by J.O. Urmson, in *The Complete Works of Aristotle*, Volume II, Jonathan Barnes, ed. (Princeton: University Press, 1984), 1140a26–29.

Judgments about the character traits or value commitments that are manifested in occupying a particular perspective, must at some point be grounded in a reflective conception of a good life. Nevertheless, making the judgment that in some particular circumstances one is being self-absorbed or uncompassionate does not require taking up a reflective point of view from which one understands the justification for thinking that these are vices and their place in a larger system of values commitments. Given these facts about normative judgments (and judgments about the character traits manifested by occupying a practical perspective in particular), we can see that recognizing reasons to shift perspectives does not require any explicit attention to one's reflective conception of a good life. Rather, we make judgments about such reasons rather automatically, without the aid of critical reflection. Granted, these "automatic" or impressionistic judgments about reasons would not have much force if they were entirely divorced from reflection and justification, however, this is not the case. Intuitive judgments about reasons are tied to reflection and justification, albeit indirectly.

Finally, one might be concerned about the fact that the aspect of wisdom I have described does not seem to involve any rational process. If we have a picture of practical rationality according to which reasoning proceeds from premises about values, goals, or desires via practical syllogisms to conclusions about actions or intentions, then the noninferential aspect of practical wisdom I have described will appear to be irrational or at least nonrational. Since wisdom and practical rationality are closely related, this seems undesirable.

The first thing to say in response to this problem is that while practical wisdom and practical rationality are closely related, we need not take them to be one and the same. If we understand practical wisdom, as Aristotle did, as including an understanding of the good and an ability to procure it, then we might say that wisdom includes but is not limited to practical rationality. The capacity for practical reason, on this view, is a capacity to meet means to ends, to draw out the conclusions of practical syllogisms or the implications of commitments. This capacity, the capacity to follow a rational procedure, is just one of the capacities possessed by the wise.

---

[19] The claim that these unreflective judgments have some authority relies on a coherentist account of justification together with a norm of practical rationality that recommends learning from experience. An intriguing, related argument for the claim that we must trust our own judgments of self-evidence can be found in Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard, 1990), pp. 177–79.

The second strategy of response is to reduce the worry that the noninferential capacity I have described is nonrational by making an analogy to theoretical rationality. In matters of theoretical rationality we can see a need for a quasi-intuitive grasp of principles that is not itself supported by the application of an inferential process. Lewis Carroll's dialogue between the tortoise and Achilles shows that there must be rules of inference that have a different status from the premises or axioms to which the rules apply.[20] One way of putting the point is that for us to be able to reason our way to conclusions, there must be some rules of inference that we grasp without reasoning to them.[21] The fact that a noninferential capacity to grasp principles is necessary for theoretical rationality should make us more comfortable with the idea that a noninferential capacity to grasp reasons is a part of practical wisdom.

This picture of practical wisdom as including a noninferential capacity stands in contrast to conceptions of practical rationality according to which rational choices include only the choices made in explicit accordance with a life plan or a coherent system of values. The practically wise person does need to have a larger perspective that encompasses all her values. She needs some conception of the human good that includes all the subsidiary goods, virtues, and vices in order to have a basis for judgment that sticking with one perspective would be insensitive, narrow-minded, self-indulgent, or the like. But many of the choices that require wisdom are not inferences from this ideal conception. Having a reflective conception of a good human life is one part of wisdom, but another important part of wisdom is being able to set this conception aside, to allow one's reflection to be informed by experience and at the same time to maintain the capacity to see one's weakness or vice from within the experience. Furthermore, given the impossibility of occupying numerous practical perspectives at once, and given the importance of being open to the lessons of experience, a reflective conception of a good life is unlikely to be the kind of detailed plan or map that would allow direct inferences about what to do in any circumstances.

---

[20] "What the Tortoise Said to Achilles," *Mind*, IV, 14 (April 1895): 278–80.

[21] Similarly, Goodman's paradox might be taken to show that we must accept or endorse the principle of induction in order to get off the ground, epistemologically. I thank Christopher Hookaway for drawing my attention to this point and Roger Crisp for helpful discussion of the Lewis Carroll example.

The preceding discussion also points to an advantage of my account of practical wisdom which is that it makes wisdom more accessible than some critics have charged. Julia Driver, for instance, accuses Aristotelian virtue ethics of being overly intellectual in its emphasis on practical reason and wisdom.[22] If practical wisdom is as I have described it, then, there is at least one important part of it that is importantly unlike a technical skill or highly intellectual capacity. Instead, practical wisdom's intellectual demands on us are like those of the other virtues: it requires that we make judgments (about character or values) that are informed by a conception of what is good for a person, but it does not always require (in fact, it may sometimes preclude) a full and reflective knowledge of this conception while we are making such judgments.

V. CONCLUSION

The person with practical wisdom should have a conception of the good that guides her in reflective moments. But this conception of the good is also shaped by the person's engagement with the world, the experience and practice that come from occupying practical perspectives. An important part of wisdom, then, is the capacity to use what we learn from experience and to judge when our being engaged in a particular way manifests a problem with our character or with our ability to pursue other values. Such judgments are, in turn, informed by a reflective conception of how to live, but the wise person makes these judgments without directly appealing to her ideal.

Pluralism about the good life is a compelling view, but we do not yet have a clear understanding of how we pursue a plurality of values in the same life in a reasonable or reflective way. The excellent but relatively small body of work on practical wisdom tends to emphasize the perceptual capacities that are involved in practical wisdom in order to distinguish it from alternative conceptions of practical rationality.[23]

---

[22] In her *Uneasy Virtue* (New York: Cambridge, 2001).

[23] For examples of this work see note 11. Much of the work in the recent resurgence of virtue ethics focuses on articulating the nature of an ethical theory that takes virtue to be the central concept rather than on providing individual characterizations of virtues. See, for example, Rosalind Hursthouse's *On Virtue Ethics* (New York: Oxford, 1999), and Christine Swanton's *Virtue Ethics: A Pluralistic View* (New York: Oxford, 2003). This is not a criticism of these virtue ethical theories; articulating a vision for virtue ethics is a vital task. If virtue ethics is to progress, however, the details of a virtue as important as practical wisdom must be given and this essay is one step on that path. One philosopher who has also taken a step along this path

There is a good reason for this as Aristotelians writing about practical wisdom have had to distinguish and defend their conception of wisdom from favored instrumental accounts of practical reason and from conceptions of practical reason that imply the commensuration of values within a single metric. We are now at a stage in the dialectic at which we can turn our attention to characterizing wise choices in the context of pluralism in more detail without having to begin from a defensive posture. I hope to have demonstrated that one way to do this is to illuminate the nature of practical wisdom by investigating some neglected details of the nature of a good life.

VALERIE TIBERIUS

University of Minnesota

(albeit in a different direction from mine) is John Kekes who gives a comprehensive account of what he calls moral wisdom in his *Moral Wisdom and Good Lives* (Ithaca: Cornell, 1995).

THE DIF*

Our topic is the difference variously labeled as that between, on the one hand, "doing" or "causing" or "making" something happen and, on the other, "allowing" or "letting" it happen. The reality and importance of this difference—which we will refer to as *the Dif*—are recognized in law and morality, it being nearly universally acknowledged that causing harm is *ceteris paribus* worse than allowing harm. Nevertheless, philosophers have found the analysis of the difference elusive and the ensuing controversy has led many to skepticism about the coherence of the distinction and/or its moral significance.

Our goal here is to say "What's the Dif?"

We take this analytical job to be the starting place for a theory of *agent* or, more generally, *object causation*. We invoke this sort of causation whenever we speak of a person or an object or a kind of thing as a cause; and we do so all the time—when we say that the moon causes the tides but not lunacy, that Vesuvius destroyed Pompeii, that the rhinovirus causes colds, that Oswald killed Kennedy. Such talk is vitally important; as when we ask what virus causes a disease, or which person is the murderer, or who let the dogs out. And it is in this idiom that we invoke the Dif: the secret service allowed, but did not cause Kennedy's death; the Federal Reserve did not prevent the recession, but it did not make it happen; by failing to destroy the rhinovirus, the immune system allows us to catch cold, but the common cold is not an autoimmune disease.

Of course, objects enter into causal relations only by doing things—that is, by participating in events—which have causal upshots. But we need a separate account of object causation because the causal consequences of what an object or agent does are not always things that the *agent* or *object* causes. That is, we do not describe an object, *o*, as a cause of some outcome *merely* because *o* does something that is a cause of that outcome. A fire alarm's failing to go off can be a cause of the destruction of a building but there is nevertheless a difference between the defective fire alarm that fails to sound when a fire breaks out and the defective fire alarm that short-circuits and

---

* We thank Jonathan Bennett, David Braybrooke, Steve Finlay, Matt Hanser, Mark Heller, and Jonathan Schaffer for their helpful comments on earlier versions of this paper. We also thank audiences at the Southern California Philosophy Conference and at the Philosophy Department of the University of California, Santa Barbara.

0022-362X/05/0204/183–205

causes the fire. It is the same kind of difference as between a lazy night watchman, liable to sleep through a fire, and a malicious watchman bent on arson. One object allows; the other causes the bad outcome.

In order either to cause or allow an outcome, something that the agent does must be a cause of the outcome. Agents cause outcomes by doing things. These doings, also called "actions," are causal *relata*; they are events in which the agent figures in the way that objects figure in events.[1] The ontology of actions is no more or less mysterious than the ontology of events generally, and while that might be very mysterious indeed, we think that nothing in our account trades upon or compounds those mysteries.[2]

Able pushes Vic into the lake. Though he could, Baker does not throw Vic a life preserver. Vic drowns. Able does something that causes Vic's death and so does Baker. Baker's refraining from throwing a life preserver is something Baker does. We see no reason not to call it an action nor do we see any reason to deny that this action is a cause of Vic's death.

Had Baker thrown the life preserver, Vic would not have died. Able causes Vic's death. But while it is correct to say that *something that Baker does* is a cause of Vic's death, we do not think it right to say that *Baker* causes Vic's death.

The difference between Able and Baker is a matter of the *way* they do the things that cause death. The *way* in which someone does something is a matter of *how* they do it.

Questions about how things are done are answered using the 'by'-locution.[3] We say "*A* does *X* by doing *Y*." The analysis of 'by' statements has been the subject of much heated controversy.[4] We are not interested here in entering this fray. Any plausible analysis of 'by' talk must respect common sense and common usage and any such analysis will, we believe, sustain our analysis of the Dif. We assume only this much, and it is an assumption shared by every analysis of the 'by' relation of which we are aware.

---

[1] Our account of the Dif could be readily recast in the language of fact causation, however.

[2] Nothing in our account will turn on any special assumptions about the nature of events or actions. There is nothing up our metaphysical sleeves. In particular we do not frontload our understanding of event causation in some way which excludes omissions or background conditions as causes. Neither do we assume any distinction between causally necessary conditions and "the" cause of an outcome. We take a maximally liberal construal of what count as actions and events. Thus we count the striking of a match as an action/event and a cause of its lighting but would also count the match's being dry as an event and a cause and the agent's refraining from getting the match wet as an action and a cause. In this we part company from those who would seek to ground the Dif in the fundamental ontology of events or the theory of event causation. Cf. Judith Jarvis Thomson, "Causation: Omissions," *Philosophy and Phenomenological Research*, LXVI (2003): 81–103.

[3] The "'by'- locution" locution is owed to Jonathan Bennett, *The Act Itself* (New York: Oxford, 1995).

[4] For an extended discussion of 'by' see Bennett, *Events and Their Names* (New York: Oxford 1988), chapters 12–14, and *The Act Itself*, chapter 2.

> *A* does *X* by doing *Y* only if:
> (i) *A* does *X*; and
> (ii) *A* does *Y*; and
> (iii) (ii) is an Insufficient but Necessary part of a condition which is logically, nomologically, or causally Sufficient for (i).

The answer, when there is one, to the question "How did *A* do *X*?" will take the form "by doing *Y*" and it will always be on the cards to ask, in turn, "How did *A* do *Y*?" and so on. *Ad infinitum*? No. It seems, typically at least, that these 'by' chains come to an end when they arrive at what some call "basic actions." Basic actions are doings for which there is no ampliative answer to the question "How did *A* do it?" Thus, we might say that someone caused the wreck by signaling and signaled by raising his hand, but we may have no answer to "How did he raise his hand?," no way of filling out, "he raised his hand by...." "He just did it" we are wont to say.[5]

Whether or not these 'by' chains must *always* come to an end is a question we shall not address here, but whether or not they do, we may say that the whole story about the *way* in which someone performs an action is a matter of the whole story about how they do what they do as revealed by these chains of 'by' dependence. So if Able and Baker both do *X* and both do *X* by doing *Y*... they may still do *X* in different ways if one does *Y* by doing *Z* and the other does not.

Sometimes 'by' relates a *causally* INS condition. Thus:

(1) He killed him by pushing him in the lake.
(2) She put out the fire by dousing him with water.
(3) He rang the bell by pushing the button.

But not all 'by's are causal. Consider:

(4) He broke his promise by not writing.
(5) She made it unanimous by voting "aye."

In cases like these the unstated conditions will complete a condition which is *logically* sufficient for the outcome: he had promised to write and he did not, she voted aye and so did everyone else.

---

[5] Though we might be able to imagine circumstances in which someone raised his hand by doing something else. Cf. Bennett, *The Act Itself*, p. 37.

There is another sort of noncausal 'by' that will be important to what follows. Consider the question "How did Baker fail to throw the life jacket?" The answer may be important. If Baker tried to throw the jacket but failed by slipping on the wet dock his culpability might be diminished. Suppose that, in our case, the answer is:

(6) Baker failed to throw the life jacket by standing on the dock, twiddling his thumbs.

Here the unstated conditions would include additional facts about Baker's behavior and environment—he did not do anything tricky with his feet, there were no mechanisms in his surround to convert thumb twiddling into life jacket throwing—facts that, together with the laws of nature, entail that Baker failed to throw the life jacket. Which is to say that if (6) is true, it will be because Baker's thumb twiddling is an insufficient but necessary part of a *nomologically sufficient* condition for Baker's failure to throw the life jacket. Even so, (6) is not causal. Though it is nomologically INS, the thumb twiddling is not a cause of Baker's failure. Had Baker not twiddled his thumbs, the jacket might still not have been thrown. He might have put his thumbs in his ears, or dealt cards, or walked away ... and the failure would still have occurred.[6]

The whole story about how someone performs an action will be told by chains of 'by' statements and the whole story might be causal. Thus it might be that *A* does *X* by doing *Y*, where doing *Y* is a cause of *X*, and that he does *Y* by doing something that causes *Y* and so on, *ad infinitum,* or until we arrive at basic actions.

In this sort of case the 'by'-chain is just a causal chain and, insofar as causation is transitive,[7] each link in the chain will be a cause of *X* and of the outcome *X* causes. Interestingly, this can be so even if each link in the 'by'-chain is not a cause of the preceding one. Thus: Swift won the medal. How? By breaking the record. How? By running a three minute mile. How? By sprinting for the last two minutes. The connection between running the three minute mile and breaking the record is logical, not causal. That he ran the race in three minutes, together with the fact that the preceding record was four minutes, entails that the record was broken. And we might have similar qualms about saying that his two minute sprint is a cause of his three minute race. The relation here, we might want to say, is that of part to whole, not cause and effect. Be that as it may, note that each link of this chain—the record breaking, the three minute race, the two minute sprint—can be properly described as a cause of Swift's winning the medal, even if it does not cause the doing that immediately precedes it in the chain.

---

[6] Nomologically INS conditions that are not causal are the bane of attempts to analyze causation in terms of laws. Notoriously, the barometer's fall is INS for the thunderstorm, but does not cause it.

[7] Note, however, that our account does not rely on the assumption that causation is transitive. For a discussion of arguments against the claim that causation is transitive, see Ned Hall, "Causation and the Price of Transitivity," this JOURNAL, XCVII, 4 (April 2000): 198–222.

In this sort of case and in the case where the 'by'-chain is a simple causal chain, each link in the chain of answers to the question "How did *A* do *X*?" reveals a cause of *X*. Let us call these ways of doing things—where causation reaches all the way down—*wholly causal* or *causal* for short.

Not every way of doing things is causal. Sometimes when we fill out the answer to the question "How did *A* do *X*?" we arrive at something *A* does that is not a cause of *X*. As we have just observed, the whole story about Baker's role in Vic's death turns noncausal when it gets to (6). The thumb twiddling and the failing to throw the life jacket are both things that Baker does. He does the latter by doing the former. But the thumb twiddling does not cause Vic's death. And while we could go on to ask "How does he twiddle his thumbs?" the answers we get will not be causes of Vic's death.[8]

In this sort of case, the explanation of how *A* does *X* runs to doings that are not themselves causes of *X*. Let us call cases like these—where the causation does not go all the way down—*noncausal*.

The difference between an agent's doing something in a causal as opposed to a non-causal way is, we maintain, precisely the difference between an agent's causing something as opposed to allowing it to happen.[9] Assuming that, in the final 'by' analysis, agents do everything they do by performing basic actions, we could say:

---

[8] Some folk have complained that it feels odd to ask "How did he do *X*?" when *X* is an omission. We agree. The actions folk call omissions are, we claim, precisely the ones that are done noncausally. Folk rarely bother to ask how omissions are performed because they recognize that the answer is not going to reveal a cause and causes are usually what folk are interested in.

[9] While we will sometimes speak of "*the* way" an agent does *X*, we do not deny that, for any *X*, there may be many different legitimate 'by' stories one can tell. One might say, with equal truth, "He bought the house by cashing in his life savings" or "by submitting a higher bid" or "by duping the appraiser." Each may be true and each begins a 'by' chain which will comprise different actions on anyone's account. We conjecture that if any of these stories are causal in our sense then any other legitimate 'by' story about how the agent does *X* must be causal as well. But this conjecture is no part of our account of the Dif. On our account an agent causes *X* if there is *a* causal 'by' story about how he brings about *X*.

(7) An agent causes an outcome *iff* the outcome is caused by a basic action of the agent.

Without appealing to the idea of basic actions, we can say:

(8) An agent *A* causes an outcome O *iff A* does something, *X*, that causes *O* and *A* does *X* causally.

Take another case. A runaway train is hurtling down the track towards an unsuspecting group of cub scouts. A bystander sees that if she throws the switch the train will be diverted to another track, where only one person, Joe, will be killed. Throwing the switch will cause Joe to die. Refraining from throwing the switch will cause the cub scouts to die. Whatever the bystander does, *something she does* will be a cause of the death of someone. But that is not to say that the *bystander* will cause a death no matter what. There is no way for the bystander to throw the switch except causally. So if she throws the switch, she will cause Joe's death. She will also cause the cub scouts not to die; that is, she will prevent their death. On the other hand, there is no way for the bystander to refrain except noncausally. So, by refraining, she will not have caused the death of the cub scouts. Thus: if the bystander refrains from throwing the switch by remaining perfectly still, it will nevertheless be false that her refraining still causes the scouts' dying. She could have refrained without remaining perfectly still, and the cubs would still have died. Likewise, while the bystander's refraining from throwing the switch is a cause of Joe's not dying, our analysis does not require us to say that in that case the *bystander* would have caused Joe not to die. More colloquially, we need not praise the bystander for preventing Joe's death but can say, instead, that she only allowed him to live.

We have defined the difference between causing and allowing in causal terms but in many of our examples we have used counterfactual dependence as a test of causal connectedness. We have taken the fact that an outcome might have happened even if the agent had not done *X* as evidence that doing *X* was not a cause of the outcome. But counterfactual dependence is not an infallible guide to causal dependence. Notoriously it fails in cases of causal preemption. Here is an example due to Frances Howard-Snyder.[10]

> An assassin, A. Sassan, is preparing to assassinate Victor by shooting him. A second assassin, Baxter, is waiting across the street watching Sassan

---

[10] "Doing vs. Allowing," in Edward N. Zalta, ed., *Stanford Encyclopedia of Philosophy* (Summer 2002 Edition on the internet).

> to ensure his success. If Sassan shows any signs of hesitation, Baxter will shoot Victor himself. Suppose Sassan knows about Baxter and his intentions and also knows that he can turn his gun on Baxter instead of on Victor if he so chooses. Although this thought crosses his mind, he quickly suppresses it, since he is committed to Victor's annihilation. He shoots Victor and Victor dies instantly.

Here it is false that if Sassan had not shot, Victor would not have died. But surely, Howard-Snyder insists, Sassan caused Victor's death by shooting. We agree. The case is one of preemption and in such cases causal dependence does not pair with counterfactual dependence. Victor's death is caused by Sassan's shooting, even if it does not counterfactually depend on it. That is why cases of preemption are standard counterexamples to the thesis that causal dependence is simply counterfactual dependence.

Even so, given our reliance on causation, it might be held that we have a special obligation in such cases to say why the dependence is causal when it is not counterfactual. While it is not our business here to endorse an analysis of event causation it seems to us that David Lewis has the right answer.[11] Roughly, Lewis proposes that to say that c causes e is not to say not that $\sim c > \sim e$, but rather that the timing and manner of e's occurrence (or the occurrence of an event like e) is counterfactually dependent on the timing and manner of c's (or an event like c's) occurrence. Lewis calls this complex relation "influence" and it may be that c influences, and hence causes, e even if $\sim c > \sim e$ is false. Thus, in the example above, while it is false that Victor would not have died if Sassan had not shot, it is true that the manner of Victor's death is highly dependent on when and how Sassan shoots. If Sassan had shot sooner or had shot differently (aiming higher or lower), then Victor would have died in a different way at a different time.

This seems to us right or at least on the right track and it underlines a phenomenon illustrated in the examples of allowing above. In noncausal cases, the way in which someone performs an allowing is usually only loosely connected with the way they move their body. There are many ways to skin a cat, but there are many, many more ways to allow a cat to be skinned.

We think that something like this observation underlies Jonathan Bennett's analysis of the Dif. Bennett draws a distinction between "making" and "allowing" in terms of an agent's alternative actions. An agent moves his body and an outcome occurs. Had the agent moved his body in a different way, the outcome would not have occurred. But now suppose that only a small fraction of the agent's range of alternative actions would have made a difference to the outcome. Had he moved in almost any other way, the outcome would still have occurred. In that case, Bennett says, the agent allowed the outcome. On the other hand, suppose the outcome would not have happened if the agent had moved in any of most of the alternative ways he could have moved. In that case, Bennett says, the agent causes the outcome. Thus, by standing by, the bystander allows the cub scouts to die because most of the alternative ways she could have moved would have the scouts dying anyway. On the other hand, had she thrown the switch, the bystander would have caused Joe's death, since most of the alternative ways she could have moved would have involved no switch throwing and no death.

---

[11] "Causation as Influence," this JOURNAL, XCVII, 4 (April 2000): 182–97.

Bennett's analysis is certainly the most successful and best argued for account heretofore. Lewis's analysis of causation as influence is useful because it can help us understand something that needs explaining: why Bennett's analysis works as well as it does over a large range of cases in capturing our intuitions about the Dif. This needs explanation because Bennett's analysans are neither obvious nor obviously connected to the moral distinction the Dif is supposed to make. As Bennett observes:

> The...distinction that I have defined *obviously* has no basic moral significance: if someone moves in a way that causes or makes probable some bad upshot, nobody would think that the moral status of his conduct depends on how many other movements by him would have done the same.[12]

This leads Bennett to conclude ultimately that the Dif makes no real moral difference.

This seems to us very wrong. Bennett's account works, when it works, because his criterion measures the extent to which the *way* someone moves his body *influences*, in Lewis's sense, whether an outcome occurs. What is really being measured is whether the agent's *body movements* cause the outcome. Thus the bystander's failure to throw the switch is a cause of the cub scouts' death. She fails to throw the switch by standing perfectly still; the fact that she could have moved in many other ways without preventing their death is relevant because it means that the way she actually moves her body is not a cause of the scouts' death. In turn, it seems to us that whether or not someone's body movements cause an outcome is *obviously* relevant to the moral status of her conduct.

---

[12] *The Act Itself*, pp. 102–03; emphasis in original.

Bennett's measure does not perfectly track influence. Howard-Snyder's Sassan case was offered as a counterexample to Bennett's analysis and in this it seems to us wholly successful. On the other hand, there is another sort of counterexample that is sometimes brought against Bennett that might also count against our own analysis. Here is Bennett's telling of the story, which we will call *Immobility*:

> Henry is in a sealed room where there is a fine metallic dust suspended in the air. If he keeps stock still for two minutes, some dust will settle in such a way as to close a tiny electric circuit which will lead to some notable upshot *U*. Thus, any movement from Henry and *U* will not obtain; perfect immobility, and we shall get *U*.[13]

Henry does not move. Bennett's analysis rules that, by remaining perfectly immobile, Henry makes *U* happen. Some find this counterintuitive. Bennett thinks the intuition can be explained away by the fact that in real life situations total immobility is rarely productive of the outcomes in which we have a quotidian interest.

What does our account say about the immobility case? The answer depends upon a prior question about the causal connection between events. Should we say that:

(9) Henry's immobility causes the dust to settle.

We are inclined to so say, but not everyone would agree. Some philosophers want, for various reasons, to deny that "omissions" or "inactions" can be causes.[14] On such views of *event* causation, (9) might be held to be false. It would follow, on our account of *agent* causation, that Henry does not cause either the settling of the dust or *U*.

In this way, the case illustrates how our theory of agent causation can serve competing intuitions about event causation. But we are suspicious of accounts that try to count out omissions as causes. We suspect that they are born of the desire to serve, at the level of event causation, intuitions that properly belong to the theory of agent causation. Indeed, we think that the failure to distinguish between agent and event causation—the failure to distinguish between what agents cause and what is caused by the things they do —is the source of much confusion in the metaphysics of causation.

---

[13] *The Act Itself*, pp. 97–98.

[14] For a succinct overview of these views of causation see Schaffer, "The Metaphysics of Causation," in *The Stanford Encyclopedia of Philosophy* (Spring 2003 Edition).

So we agree with Bennett's verdict in this case. Nevertheless, we think his account fails and fails not just because of preemption counterexamples. Consider another case of our own:

> *Mobility*: Henry is in a sealed room with an extremely sensitive motion detector. If Henry moves at all, the detector will trigger some notable upshot *U*. Only if he keeps stock still for two minutes will *U* fail to obtain. Henry moves.

On Bennett's analysis, if Henry moves, he will only have allowed *U* to occur since *U* would have happened for almost every way Henry could have moved. This seems to us obviously false. Henry's moving causes *U* by as clear a chain of cause and effect as a carom of billiard balls—never mind that he could cause the outcome by moving in any way at all. So Henry causes *U*.

Causation makes the Dif, and, it seems to us, it is a difference which manifestly bears all the moral significance which tradition and commonsense confer on it.

II

Let us consider some cases.

First, as points of moral and causal reference, we have paradigm cases of agent causing:

> *Push*: A cart stands at the top of a hill. Agent pushes it. The cart rolls down the hill and fatally injures a child.

and allowing:

> *Stayback*: The cart is already rolling; Agent could but does not interpose a rock which would stop it. The cart rolls down the hill and fatally injures a child.[15]

Any reader who does not feel a *prima facie* moral difference in these cases probably need not read on. For those who intuit a difference, however much they might suspect or disdain that intuition, let us use these cases as touchstones on the rocky road ahead.

Things quickly get tendentious when we turn to:

---

[15] The cart setup and the specific examples Push, Stayback, and Kick are due to Bennett, *The Act Itself*, p. 67.

Dislodge: A cart stands at the top of a hill, its wheel chocked by a rock. Agent kicks the rock away. The cart rolls down the hill and fatally injures a child.

Kick: The cart is rolling towards a point where there is a rock that would bring it to a halt. Agent kicks away the rock; the cart rolls down the hill and fatally injures a child.

Do the kicks in Dislodge and Kick count as causes of death? There are theories of causation that would say not. Dislodge is a case of what Jonathan Schaffer calls "causation by disconnection" and cases of this sort are counterexamples for theories that require some flow of force or energy from cause to effect.[16] Kick is an instance of the broader class of cases which the literature calls "preventing a preventer" and they will seem problematic for any theory that requires chains of spatiotemporal contiguity or persisting processes between cause and effect or any theory that wants to prohibit absences (such as, of the obstructing rock), or omissions (such as, the cart's not striking the rock) as causes.[17]

In respect to the purely causal issue, it seems to us that any theory that fails to count the kicks in Dislodge and Kick as *bona fide* causes is *prima facie* dubious. As others have pointed out, causation by disconnection, by the prevention of preventers, by omission and absence, are ubiquitous in everyday life and are there unhesitatingly counted as real causation by anyone who does not have a metaphysical axe to grind.[18]

On the other hand, with respect to the division of cases into doings and allowings, things are not so clear. There is a temptation, stronger in Kick than Dislodge but undeniably present in both, to describe these agents as having, by their actions, "allowed the cart to roll down the hill and kill the child." This way of speaking, in turn, might incline us to a theory of the Dif which pushes disconnection and preventer prevention to the allowing side of the divide.

---

[16] See, for instance, David Fair, "Causation and the Flow of Energy," *Erkenntnis*, XIV (1979): 219–50.

[17] Cf. Phil Dowe, *Physical Causation* (New York: Cambridge, 2000); David Armstrong, "Going Through the Open Door Again: Counterfactual versus Singularist Theories of Causation," in Gerhard Preyer and Frank Siebelt, eds., *Reality and Humean Supervenience* (Lanham, MD: Rowman and Littlefield, 2001), pp. 163–76; Helen Beebee, "Causing and Nothingness," in John Collins, Ned Hall, and L.A. Paul, eds., *Causation and Counterfactuals* (Cambridge: MIT, 2004), pp. 291–308.

[18] See Schaffer, "Causation by Disconnection," *Philosophy of Science*, LXVII (2000): 285–300; Lewis, "Causation," in *Philosophical Papers*, Volume 2 (New York: Oxford, 1986), pp. 159–213; H.L.A. Hart and Tony Honore, *Causation in the Law* (New York, Oxford, 1959).

The problem with this route is that the differences between preventing and enabling, acting and omitting seem as evanescent as the Dif itself. Here is Bennett on the proposal to treat Kick as an allowing because the agent only prevents a preventer:

> But then in Push, when the vehicle was standing motionless, that was because it was prevented from rolling by small stones, holes in the ground, wet grass, and so on; so an agent in pushing the vehicle was preventing these things from preventing it from moving; yet in this case we do not say that in pushing the vehicle he allowed it to move. What makes the difference? I am not sure; though it seems to involve the difference between removing obstacles (Kick) and enabling something to overcome obstacles (Push). Enough of this! We should not look for help from the meaning of 'allow' which is a complex shallow mess.[19]

Bennett's position is that Kick belongs, where his account would place it, on the making (Push) side of the Dif.

We have already committed ourselves to counting the kicks in Dislodge and Kick as causes of the cart's motion and that commits us to treating these agents as causing the child's death. In support of this way of carving things up, we may note that whatever temptation there may be to use the language of 'allowing' in these cases there is no corresponding temptation to treat these agents *morally* on a par with the agent of Stayback. In real life cases of disconnection and preventer prevention—the hangman springing the gallows' trap, the vandal who removes the guard rail at the tourist lookout or the greedy pharmacist who waters the vaccine—we count the agents as killers on a par with Push and not bystanders like Stayback.[20]

Let us acknowledge the obligation to explain away the contrary intuition here—to do something to clean up the mess about which Bennett complains—and take up a more pressing problem.

The problem is *Step Away*.

> *Step Away*: The cart is already rolling; Agent is in its path at the top of the hill and if he remained where he was the cart would bump into him harmlessly and stop. Agent realizes this but steps out of the way and the cart rolls down the hill and fatally injures a child.

It seems that the only difference between Step Away and Kick is that in the former case it is the agent's body, not the rock, which is the obstacle to the cart. But this seems not much of a difference. Indeed, we can narrow the difference more by imagining that in Step Away the agent steps away by kicking a rock by the side of the path, so that the agent's body moves precisely as it does in Kick and, of course, with the same upshot—the death of the child.

---

[19] *The Act Itself*, p. 68.

[20] It is easy to get derailed by features of examples that seem trivial. One's first intuition is that the agent of Push is more likely to be counted guilty of murder in the first degree than the agent of Kick but that is only because, in the normal course of things, kicking away is not a likely technique for premeditated murder.

This argues for treating Step Away as a killing on a par with Kick. The problem is that Step Away looks like an instance of *ducking.*

> *Duck*: Agent sees that the bad guy is about to shoot in his direction. Agent ducks and the shot kills a child.

Ducking was introduced to the literature by Christopher Boorse and Roy Sorensen.[21] They contrasted ducking with *sacrificing.*

> *Sacrifice*: Agent sees that the bad guy is about to shoot in his direction. Agent grabs a child and interposes its body between himself and the gunman. The bad guy shoots and the child is killed.

Boorse and Sorensen thought it clear that the agent of Duck is no killer, while the agent of Sacrifice is. If we agree with that, then it seems that we ought to treat Step Away as a case of ducking and we are stuck with explaining the difference between Kick and Step Away.

Bennett's account offers a way to do so. Plausibly, of all the ways of moving available to the Kick agent, only a few constitute a kicking of the rock. On the other hand we can imagine that there are many different ways that the Step Away agent might move, any one of which would be a stepping (or hopping or jumping ...) away. This would allow Bennett to treat Step Away as an allowing and Kick as a killing.

The trouble with this is that we can also imagine that there are *not* a lot of alternative ways of stepping away. Jeff McMahan[22] describes this case:

> *Lunge Away*: Like Step Away except that Agent can only remove himself from the path of the cart by hastily executing a precise set of specific body movements.

Bennett must count this as a killing and of course will have to precisely parse the continuum of cases where varying restrictions on range of motion transform a Step Away agent's *allowing* to a

---

[21] "Ducking Harm," this JOURNAL, LXXXV, 3 (March 1988): 115–34.

[22] "A Challenge to Common Sense Morality," *Ethics,* CVIII (1998): 394–418, see p. 404.

Lunge Away agent's *killing*. And if this seems arbitrary, Bennett can always reply that that is why he thinks it an illusion that *the Dif* morally matters.

One begins to see why Boorse and Sorensen introduced making sense of ducking and sacrificing cases as "an adequacy condition on solutions to the act/omission problem."[23] Interestingly, though, they did not think that the Ducking-Sacrificing divide is equivalent to the Doing-Allowing Dif. Their evidence was a case we can call *Drive Away*:

> *Drive Away*: Agent's car is parked between a rolling cart and a child. Agent sees the child in the cart's path but drives off to avoid an ugly dent.

They argued that if Duck is an allowing then so is Drive Away. But, they insisted, Drive Away, unlike Duck, is a case of "culpable homicide." The latter agent "... can be convicted of murder or manslaughter in any American court. On the other hand, had [Agent] merely failed to interpose his car between rock and child, he would be innocent in both Anglo-American law and, presumably, an ethics that takes the act/omission distinction seriously" (*ibid.*, p. 127).

Considering this case, Alison McIntyre concludes that since the *only* difference between Drive Away and Duck is a moral one this shows that our apparently causal judgments are sometimes fundamentally controlled by our moral assumptions about what one is obliged to do:

> Since the cases are similar except in their moral characteristics, what lies behind this judgment must be the assumption that there is a moral duty to risk damage to your car to protect the child, and a consequent expectation that you will do so but that there is no duty for you to risk injury to yourself—hence, moving your body out of the way does not cause the child to be injured and is not a case of causing the child's death.[24]

But it would be very remarkable indeed if our *moral* assumptions were to drive our *causal* judgments in these cases. Consider that, notwithstanding the expectation that one need not sacrifice oneself, we are not in the least tempted to deny that the agent who shoots an assailant in self defense *causes* the death of his assailant.

Shifting perspective, McMahan offers another sort of case that suggests that our intuitions on *any* of these cases might be altered simply by changing its surround:

---

[23] Boorse and Sorensen, p. 115.

[24] "Guilty Bystanders? On the Legitimacy of Duty to Rescue Statutes," *Philosophy and Public Affairs*, XXIII (1994): 169–70.

> *Interpose Then Kick*: The cart is already rolling. Agent places a rock in the path of the cart. If left in place, the rock would stop the cart; but before the rock can stop the cart, Agent reconsiders and kicks the rock away. The cart rolls down the hill and fatally injures a child.

McMahan argues that even if Kick is killing, Interpose Then Kick looks more like allowing.[25] After all, if we call this killing what should we say about *Walk By*?

> *Walk By*: The cart is rolling down the hill. Agent walks into its path. If he remained where he is, the cart would bump harmlessly into Agent and stop. But before the cart reaches him, Agent walks on.

And how, in turn, should we distinguish Walk By from Step Away?

Interpose Then Kick and Walk By are only two of a whole genre of cases that McMahan dubs "withdrawing aid."[26] Other examples:

> *Dutch Boy*: Dutch Boy spots a leak in the dyke. The leak threatens to ultimately wear away at the structure, collapse the dyke, and destroy the village. Dutch Boy plugs the leak with his finger but, after a time, tires of his situation and pulls his finger out. The leak resumes, destroys the dyke, and wreaks havoc on the village.

> *Involuntary Donor*: Agent finds himself connected by a medical apparatus to a sickly violinist. The apparatus enables Agent's body to sustain the violinist. Tiring of his situation, Agent unhooks himself from the apparatus. The violinist dies.

Withdrawing aid, McMahan argued, is so different from either killing or allowing as to require an entirely separate status.

So what do *we* say about all these cases? Which are killings, which are lettings die? A first answer: in some cases, both.

Traditionally, accounts of the Dif have been *summary*, in the sense that taking everything the agent does into account they judge that he has either caused or allowed an outcome. This is not so with our account; an agent may cause a death by doing one thing and allow it by doing another. Able caused Vic's death by pushing him in the lake and allowed him to drown by failing to throw him a life jacket.

---

[25] "This background detail inclines us to locate this kick on the negative side of the generic positive-negative divide"—McMahan, "A Challenge to Common Sense Morality," p. 402.

[26] McMahan, "Killing, Letting Die, and Withdrawing Aid," *Ethics*, CIII (January 1993): 250–79.

Many of the cases just surveyed are bivalent in this way. Thus, after the Dutch Boy has removed his finger it is easy to imagine that he could reinsert it and save the day. So we should say that by failing to put his finger back in the Dutch Boy allows the ruin of the village even if we also say that he causes it by pulling his finger out in the first place. Likewise we could imagine that the agents in Involuntary Donor and Step Away might have had second thoughts. The Involuntary Donor could reconnect the machine. The agent in Step Away could step back. By failing to repent, these agents allow the harmful outcomes they cause.

That one can allow as well as cause explains our temptation in cases like Kick to describe the agent as "allowing the cart to roll down the hill" since that is precisely what the agent does in the interval before and after he kicks the stone away. If the interval of allowing is short, as it will typically be in a case of Dislodge, there is less temptation to use 'allowing' talk. And where little or no allowing is likely, as in Push, there is little or no temptation to describe what has gone on as "allowing the cart to roll."[27]

More careful story telling can remove the dissonance. Suppose that the damn bursts the instant the finger is withdrawn; that the patient dies the moment the machine is disconnected; imagine that the stone cannot be reinterposed. Thus redescribed, we think the temptation to see these as allowings recedes.

So, setting aside the distraction that some of these cases may be allowings as well as killings, which ones are also killings?

Let us put our cards on the table. We think that with the exception of Stayback, *all* agents in the cases we have considered in this section—Push, Dislodge, Kick, Step Away, Duck, Sacrifice, Lunge Away, Drive Away, Interpose Then Kick, Walk By, Dutch Boy, Involuntary Donor—are killers; each causes, not merely allows, death.

Let us quickly add, however, that we do *not* hold that these cases are *morally* equivalent. Indeed, we offer it as a central argument for our account of the Dif that it can make sense of the moral *differences* among these cases.

III

The Oxford English Dictionary defines consequentialism as "an ethical doctrine which holds that the morality of an action is to be judged solely by its consequences." But, of course, it is not so. The death of an innocent is a bad consequence, but no consequentialist would condemn an action just because it has this upshot. It depends on the alternatives. If the agent's choice is between killing one innocent and killing many, the consequentialist would recommend killing one. A consequentialist wants to say things like "it is better to kill one innocent than two," but, of course what this means is that you should kill one rather than two *if those are your alternatives.* It does not mean that if Able kills one innocent and Baker kills two then Able's act must be better than Baker's. It depends on the alternatives. Better to say, "other things being equal, one killing is better than two," where what are equal are the *moral* consequences of the alternatives to the killings.

---

[27] We should note, but will not here elaborate upon the other complication here; namely, that as an upshot of the agent's actions the *rock* allows, where it would otherwise have prevented, the cart to roll down the hill.

A deontologist is someone who thinks that *how* an action is performed is morally significant. The deontological commonplace with which we began was that it is worse to kill someone than to allow someone to die. But, come to think of it, that cannot be the assumption which has been animating our intuitions about these cases. Taken at face value, that principle would recommend allowing someone to die over killing them. Given such a choice this might or might not be sound advice, but that is *not* the sort of case we have been considering.

Our paradigmatic allower, Stayback, does not choose between allowing the child to die and killing him. By staying back he allows the child to die rather than *preventing* his death. And our paradigm killer, Push, has not chosen to kill the child rather than allow him to die; he has chosen to kill over allowing the child to *live.*

Nor does it help to qualify by saying "other things being equal it is worse to kill than allow." For what is being held equal in our judgment that Push is worse than Stayback? One might think the alternatives are morally equivalent since, in both, a child does not die. But that would be to judge the alternatives by their *consequences.* The alternative *actions* in these cases are allowing a child to live and saving a child's life and the same deontological intuitions that tell us that *causing* is worse than *allowing* the death of a child also cry out that *saving* a child's life is *better* than *allowing* it to live. If the agent of Stayback had not stayed back we would have praised him, but we do not praise people for refraining from killing children.

So the principle *really* at work in our judgment that Push is worse than Stayback seems to be:

> Killing someone, when the alternative is to allow that person to live, is worse than allowing someone to die, when the alternative is to prevent that person's death.

Now all of this might not be worth saying if the alternatives were always the same. But are they? Must it always be that the alternative to killing is to allow life to continue, that the alternative to allowing a death is saving a life? More generally, must it always be that an agent's only alternative to causing an outcome will be to allow it not to happen?

The answer will depend on one's theory of the difference between doing and allowing. On our account, the answer is "no."

To see this, consider the difference between Kick and Step Away. In both of these cases a child dies as a causal upshot of an agent's basic action. In one case it is a kick; in the other, a step. As we observed above, the choreography of these actions may be precisely identical. On our account, both agents cause a child's death. But there is a difference.

Consider the situation if the agent of Kick had not kicked away the rock. His refraining from kicking the rock would have been a cause of the child's survival. How might he have refrained? Well, by stepping back, or dancing a two-step, or sitting down.... It does not much matter. Refraining from kicking a rock is no basic action; to perform it the agent would have to do some action or other (other than kick that is). Any action that, in the circumstances, nomologically suffices for the rock not moving would serve. But nothing that the agent does in this vein will be a *cause* of the child's survival. This is to say, if the agent of Kick had not kicked the rock away he would only have allowed the child to live; he would not have saved the child's life.

Compare this to the case of Step Away. Suppose that agent had not stepped away; that he had, instead, remained in place. How might he have done that? Well, he might have stepped briefly to the side then back again. Or he might have lunged out of the way and then crawled back into the cart's path. Or the agent might have performed that most basic of basic actions: he might have just stood still. In any of these alternatives, the cart would be stopped by a basic action of the agent. So, by our reckoning, the agent would have caused the child not to die, or in other words, prevented the child's death.

This, in turn, is why, had the Step Away agent not stepped away, we would praise him for saving the child's life. If the bump he endured from stopping the cart were great, we would praise him all the more. On the other hand, if the Kick agent had just stayed back he would earn no praise. He would only have allowed the child to live. In so doing he would have obeyed morality's minimal requirements but no hero he.

The Kick agent chooses to cause the child's death rather than allow it to live. The Step Away agent chooses to cause the child's death rather than prevent its death.

It is this difference in the moral weight of the *alternative* actions open to them that explains, we think, the moral differences between the agents with respect to the actions they *actually* take. Both agents cause a child's death, but the agent of Step Away is morally on a par with the agent of Stayback, who only allows a death, while Kick looks every bit as bad as Push.

Generalizing this observation leads us to what we shall call the *Deontological Equivalence Principle.* The principle might be put like this:

> *Deontological Equivalence*: Causing a harm, when the alternative is preventing that harm, is morally equivalent to allowing a harm when the alternative is preventing that harm.

Where the harm is death:

> Causing a person's death, when the alternative is preventing that person's death, is morally equivalent to allowing a person to die, when the alternative is preventing that death.

More colloquially:

> Causing a death you would otherwise be preventing is no worse than allowing it.

On this principle, the agents of Step Away, Duck, Lunge Away, Walk By, Dutch Boy, and Involuntary Donor, though they all cause death, are morally no worse than Stayback; no worse, that is, than the person who allows someone to die rather than save him.

The agents in these cases cause death by refraining from preventing it. An ordinary killer of the Push or Kick sort causes death by refraining from allowing the victim to live. The Equivalence Principle allows us to treat an agent as causing a death without having to regard him as morally on a par with an ordinary killer. And separating the moral from the causal judgments allows us to accommodate our intuitions about the differences in these cases without having to gerrymander our account of the Dif.

To understand the difference between our account and mere gerrymandering it is important to understand what we are *not* arguing when we advocate the Equivalence Principle.

Looking at these cases, one might think that what they have in common is, roughly speaking, that the benefits that the agents deny victims are ones that the agents themselves have conferred. Thus, the barriers to harm that Dutch Boy and Involuntary Donor remove are the agent's own body. Accordingly, one might argue that in such cases the agents should be cut moral slack because, as it were, they are only taking away with one hand a good given with the other.[28] This is not our argument.

Nor are we invoking the idea, often voiced in discussions of the Dif, that in cases like these, by failing to prevent deaths (albeit by causing them), the agents are merely "allowing nature to take its course."[29]

Nor are we trying to argue that these cases are special because in them the final upshots of what the agent does are the same as if the agent had never existed.[30]

All of these ways of arguing try to justify our deontological intuitions by appeal to some sort of summation of the values of the *consequences* of an agent's act or actions. Observe that following any or all of these lines of thought to their logical conclusion would require us to treat Interpose Then Kick as on a par with Step Away and the rest. After all, in Interpose Then Kick, the agent only removes a benefit—a barrier to the cart —that he himself has provided. He allows nature, in the form of the already careening cart, to take its course. And once the stone is kicked away, for all the difference the agent has made, he might never have been there at all.

From our point of view this way of thinking is deontology manqué. The Equivalence principle does not apply to Interpose Then Kick. How the stone got there, what would have happened had the agent never existed...are beside the *deontological* point. The relevant deontological facts are those about how the agent acted. The agent was there and kicked the stone. Had he refrained from kicking, he would have allowed the child to live. We think he is a killer, no better than the agent in Push and we think that common sense would agree. Courts would convict him as readily as the agent of Drive Away.

But now what of Drive Away? Why does intuition treat this case as so different from Duck?

Consider how the agent acts. The agent moves his car out of the way. If it is an ordinary sort of car—and that is surely what our ready judgment assumes—then it must be that the agent makes the car move either by releasing the brake or pressing the throttle or both. In the case of ordinary cars only some such basic actions will get the things going. Suppose, in this case, the agent causes the car to move, and hence the child's death, by stepping on the throttle.

---

[28] Cf. McMahan, "Killing, Letting Die, and Withdrawing Aid."

[29] Cf. Alan Donagan's account in *The Theory of Morality* (Chicago: University Press, 1977); and Philippa Foot, "Killing and Letting Die," in Bonnie Steinbock and Alastair Norcross, eds., *Killing and Letting Die* (New York: Fordham 1994, second edition), pp. 280–89.

[30] Shelly Kagan, *The Limits of Morality* (New York: Oxford, 1989), pp. 94–96.

The agent's alternative was not to move the car. How could he have accomplished that? Well, by not stepping on the throttle. But not-stepping-on-the-throttle is no basic action and, however the agent performed it, no basic act of his would have been a cause of the car's motion or the child's death. So the driver's choice was between causing a death and allowing the child to live. The Equivalence Principle does not apply. The driver is as much a killer as the agent in Push, which is why the juries would convict.

Of course, this intuition is founded on the assumption that the car is an ordinary sort of car worked in an ordinary sort of way. What if it was an *extraordinary car*, a car rigged, say, so that only the driver's total immobility would keep it from moving? If that were the set up, the driver would cause the child's death by moving and his alternative would be to prevent the child from dying by total immobility. In that case, would not the Equivalence Principle apply?

It would. And, given these extraordinary details, we think no jury would convict. We can be confident at least that Jonathan Bennett would not convict. For, as now described, the case is—automotive elaborations aside—the Mobility case we considered early on. Bennett wanted to call that case an allowing. We disagreed, but now we can see why intuitions might run Bennett's way. The agent in Mobility causes the bad outcome *O*, but because his alternative is to prevent *O*, the Equivalence Principle applies. So Mobility is morally equivalent to an allowing. On the other hand, the Equivalence Principle *also* applies to the Immobility case, which may explain why Bennett's critics balked at counting it, as Bennett wanted to count it, as plain doing.

One could accept the Deontological Equivalence principle without accepting our account of the Dif. But to give the principle any purchase requires an account of the Dif which allows cases where the alternative to causing death may not be that of allowing life. Note that Bennett's account does not allow this. On Bennett's account, an agent makes an outcome happen if it is one of a minority of the ways the agent might have moved that have the outcome as an upshot. It follows that had the agent moved in one of the majority of ways which do not have that upshot, his action would be an allowing. Which is to say that on Bennett's account an agent's alternative to making an outcome occur is inevitably allowing it not to occur. Bennett's account is not unique in this. So far as we know, every account of the Dif other than our own entails that the alternative to killing must be allowing to live.

The Equivalence Principle morally equates actions given certain alternatives; as we might expect, the equivalence holds for the alternatives as well.

> *Deontological Equivalence 2*: Preventing a harm, when the alternative is causing that harm, is morally equivalent to preventing a harm when the alternative is allowing that harm.

Which, applied to death, gives us:

> Preventing a person's death, when the alternative is causing that person's death, is morally equivalent to preventing a person's death when the alternative is allowing that person to die.

The second form of the principle deals with cases in which the agent refrains from causing a death by preventing it. For instance:

> *No Step Away*: A cart is rolling down a hill. Agent is in its path at the top of the hill and if he remains where he is the cart will bump into him and stop. Agent realizes this and remains where he is.

Before appreciating the equivalence we might have thought that causing a death must always be as impermissible as it is in Push and that refraining from causing death must always be as obligatory as:

> *No Push*: A cart stands at the top of a hill. If pushed it would roll down the hill and fatally injure a child. Agent does not push.

But the second Equivalence Principle tells us to treat the No Step Away agent as morally equivalent to the agent of No Stay Back:

> *No Stay Back*: The cart is already rolling. Agent interposes a rock which stops it before it can hit the child.

The No Step Away and No Push agents both refrain from causing a child's death but they are morally different. The No Step Away agent is laudable; if the bump he received was sharp, maybe even heroic. Like the No Stay Back agent, he is a Good Samaritan. By not pushing, the agent of No Push has likewise refrained from causing a death. He has done the right thing but he is no Good Samaritan; he has done no more than any of us do when we refrain from manslaughter.

The mention of Samaritans will likely recall Judith Jarvis Thomson's 1971 article "A Defense of Abortion."[31] Of all our examples, Involuntary Donor bears a special seriousness since Thomson invented it as an analogy for abortion. Thomson conceded that by unplugging himself the agent kills, but she did not think it right to call it murder.

---

[31] Reprinted in William Parent, ed., *Rights, Restitution, and Risk: Essays in Moral Theory* (Cambridge: Harvard, 1986), pp. 1–19.

She tried to make her case in the language of rights. She argued that the victim has a right to life, but not a right to be saved and wanted to say that by unplugging, the involuntary donor was only failing to save, not killing, or anyway not killing unjustly. The donor had only failed to be a Good Samaritan.

We have not appealed to rights or justice or samaritanism but have arrived at a consonant result. The Equivalence Principle vindicates Thomson's core intuition. When a pregnant woman acts to bring about an abortion, it is clear that the alternative to this causing of death is a prolonging of life; her action is morally equivalent to allowing a death.

This is not to say that abortion or allowing-to-die is morally blameless. For all we have said, those who think that allowing someone to die is morally equivalent to killing may continue to say so. But they are not entitled to say that they can not see the Dif.

KADRI VIHVELIN

University of Southern California

TERRANCE TOMKOW

Dalhousie University

BOOK REVIEWS

In Mendel's Mirror: Philosophical Reflections on Biology. PHILIP KITCHER. New York: Oxford University Press, 2003. xv + 385 p. Cloth $80.00, paper $35.00.

A collection of essays by a preeminent philosopher of biology, this book covers a remarkable range of topics: reduction, Darwinism, species, teleology, altruism, cultural transmission, race, genetic determinism, the Human Genome Project, sociobiology, and the "new creationism." To each of these topics, Philip Kitcher brings a characteristic mix of analytic and mathematical talent, deep empirical knowledge, boldness, iconoclasm, and in every single chapter, some considerable novelty. It is an excellent thing that these essays have been reprinted together: they not only illuminate the patterns of thought of an influential philosopher, useful even to those who have read many of the essays before, but also constitute together a sophisticated introduction to contemporary philosophy of biology.

The first four chapters develop a subtle view of inter-level relationships in biology: genetics in chapters 1 and 2 and evolutionary theory in 3 and 4. Kitcher's main thrust is toward autonomy at these levels, though in chapter 4, an essay written jointly with Kim Sterelny, he defends the possibility of describing evolution from a purely gene's-eye point of view. This last is a masterly essay and puts into question a great deal of theorizing about multiple levels of selection. Chapter 13 applies these ideas to a negative assessment of critiques by Richard Lewontin, Susan Oyama, and others of the idea that phenotypes result from gene-environment interaction. These authors argue that such interactionism implies genetic determinism; Kitcher shows decisively that they are wrong. Chapters 5 and 6 defend a "pluralistic" view of species, but this view seems out of step with chapter 11, a bold defense of the biological basis of race; chapter 11 takes the view that reproductive isolation is the *only* criterion that would make race a biological category, but chapter 5 denies that there is a single correct basis for biological taxonomy. Chapter 7 defends a unified view of function and design, surprisingly across artifacts and biological entities. Chapters 8 through 10 deal with altruism and cultural transmission.

In recent years, Kitcher has been preoccupied with social questions. In his writings on genomics (chapter 12 is an example), he chillingly

0022-362X/05/0204/206–16

demonstrates the moral relevance of mere possibility. A hundred years ago, when gene-therapies were unknown and eugenics was an application of animal breeding, one could abjure eugenics by simply not interfering in human reproductive choices. Once we come to "know how to identify the genotypes of future people," however, and a decision to abort a fetus in order to forestall some of these is founded on good science, noninterference becomes a eugenic choice because it adopts or coerces one of a number of feasible alternatives with regard to reproductive activity. Suddenly, eugenics is "the only option" (263). Thus, the choice is between good and bad "eu"-genics. In negotiating such social decisions, Kitcher urges that we have a duty to ensure that, with a few qualifications, resources are redistributed to the extent possible from those who live above a minimum acceptable standard to those who live below it: similar increments of quality of life to those above and those below this standard are not morally equivalent. These insights are not as well integrated with one another as one might have wished, but each provides material for reflection.

Chapters 14 and 15 are a pair of essays on human sociobiology which I found disappointing coming from the author of the magisterial (if sometimes overly polemical) *Vaulting Ambition.*[1] Chapter 14 advances a complicated framework for the measurement and explanation of incest-avoidance taboos. The proposal seems not to rest on empirical constraint: no argument is made for the relevance of all the parameters introduced. Though one might concede that a number of early incest studies are empirically flawed, the reader will find no strong motivation for Kitcher's own ramified account. For example, he subsumes sexual nonattraction to individuals familiar during childhood to a process of "habituation" in early development (312)—habituation being the quite general phenomenon of an organism's losing interest in stimuli that are repeatedly presented. This hypothesis is often advanced by psychologists to explain diminished sexual excitement after repeated contact with a given partner, but, as far as I can see, it has very little plausibility as an explanation of incest avoidance. First, only *sexual* contact is supposed to habituate an individual to sexual contact with a particular individual, and incest-avoidance *prevents* sexual contact. Secondly, organisms habituated to a stimulus can become interested in it again if it is presented after an interval; incest avoidance, however, is supposed to be imprinted for life. Lastly, there is little motivation for supposing that such habituation would be restricted to early development, as imprinting for incest avoidance is meant to be. Thus, habituation would militate against monogamy, yet humans are supposed to be at worst "mildly polygynous"[2]; moreover, habituation is universal across organisms, while neither incest avoidance nor monogamy is so.

---

[1] Cambridge: MIT Press, 1985

Chapter 15 criticizes E. O. Wilson's attempts to biologicize ethics. It takes Wilson at his own philosophically naïve word, and uses this to berate him in a patronizing way. Wilson believes (a) that certain deep psychological commitments to not harming, and indeed helping, others have an evolutionary origin, and (b) that this and other deep psychological commitments need to find space or even validation in socially coerced codes of behavior. He argues that these propositions have metaethical implications. Here he may well be wrong in both general thrust and fine detail, but I find (a) and (b) themselves far more interesting than any elementary metaethical fallacies they may encourage in the philosophically untutored. It is disappointing to find words like "ham-fisted" and "bungle" used of a great thinker on such flimsy grounds, and to see him accused of "lack[ing] crucial intellectual tools" and "desert[ing] standards of rigor and clarity" (331).

Chapter 16 (coauthored with A. Leah Vickers), an attack on certain "pop sociobiological" theories of female attractiveness and rape, is considerably better: here, though it might have been nice to see serious human sociobiology acknowledged at least in passing, the target amply deserves derision. Vickers and Kitcher rightly point out that it is a mistake to take rape to be a coercive variant of the same sexual preferences as are manifested in consensual relations. At this point, however, they make a wrong turn: they suppose that "the crucial question is why some men hold back from forcing women to their desires and others don't" (351). Surely, this focus on *inhibitory* mechanisms against rape is an instance of the mistake just mentioned: it assumes that all men want to rape women attractive to them, but some "hold back." But rape is not merely an instrumental choice: "I want this woman: shall I have her by securing her consent or not?" is not the question most men face. Normally, men want only consensual relations. Actually, even on the most simplistic model of rape being motivated by attraction, it is easy to understand why this act is repellent to most men. Without modern weapons used as a threat, rape is extremely violent; in evolutionary times, it would have involved the penetration of a bludgeoned, bruised and bleeding, semi-conscious or dying female—not the young girl with "bouncy gait" and "lustrous hair" so evocative of "high reproductive value" (344).

---

[2] Jared Diamond, *The Third Chimpanzee: The Evolution and Future of the Human Animal* (New York: Harper, 1993), pp. 70–71.

Vickers and Kitcher point out that "the evolutionary reasons why a young man can get an erection just by looking at a photo of a naked woman" do not automatically translate into evolutionary reasons why some young men rape. They are right. However, they do not make enough of the fact that such autonomic reactions do not necessarily translate into *desire* for sexual contact, by force if necessary, which must then be inhibited in the case of nonraping males. A human being's propensities concerning sexual contact—his or her underlying "sexual strategy"—may, after all, include aversion even in some cases where simple autonomic reactions are induced (and even women are subject to these, I am told). This, by the way, is a legitimate query for human sociobiology: Despite the nonselectiveness of arousal, humans have differentiating preferences with regard to sexual partners—Why? One might call this the metapreference problem. It is a clue to the nature of preference that though assortative mating is an important social phenomonon, leading, for example, to economic inequality because of the mutual preference of high earners, full interfertility has been maintained between human groups. Apparently, preference is cumulatively blind to those inherited characteristics that distinguish races and other groups, at least within homo sapiens.

So much for an overview of Kitcher's book. I turn now to a more detailed consideration of his views on two central issues in the philosophy of biology: reduction and species.

Inter-level Relationships in the Biological Sciences. In 1975, Hilary Putnam argued that one does not need to resort to a description in terms of micro-properties in order to understand why a rigid square peg of 1-inch side does not fit through a hole of 1-inch diameter.[3] The macro-level properties just mentioned are quite sufficient. Putnam's argument is widely thought to depend on what is known as "multiple realizability": pegs made of different materials will all fail to pass through the hole, and so the composition of the peg makes no difference to the case. Putnam himself did not think that multiple realizability was of any consequence. Speaking of the reduction of psychological properties, he says: "Even if it were not physically possible to realize human psychology in a creature made of anything but the usual protoplasm, DNA, etc.,...[the identification of psychological states with their physical realizations would still lack] *explanatory* value in

---

[3] "Philosophy and Our Mental Life," in *Mind, Language, and Reality* (New York: Cambridge, 1975), pp. 291–303.

psychology" (*ibid.*, p. 293).[4] Kitcher's argument in chapters 1–3 of the present book is an application of Putnam's peg-and-hole: it shows how the *ordinary* practice of biology is nonreductionistic in tenor, even though it takes "the usual protoplasm, DNA, etc." for granted.

The argument rests on the idea that some phenomena are not better explained when detail is uncovered. Kitcher suggests that the individual explanations proffered by a science operate autonomously at several different levels. The classical genetics of Thomas Morgan, Alfred Henry Sturtevant, and Hermann Müller rests on a certain characteristic style of reasoning:

> In certain paired entities, there is a correspondence relation between the parts of one member of the pair and the "homologous" parts of the other member. In a characteristic process, the pairs exchange homologous segments and separate. The chances that segments belonging to a single member of a pair end up together after this process of exchange is proportionate to the distance that separated them initially.

This abstract model is used to explain the outcome of genetic recombination. Commensurate with the above, there is a second style of reasoning, which calculates the distance between segments of the original pairs based on how frequently these segments assort independently. As David Hull puts it, "parents differing in some trait are crossed and the distribution of the trait in hybrid progeny is observed" (quoted by Kitcher on p. 5). In classical genetics, this method is used to construct what are sometimes known as "gene maps." Kitcher claims, rightly in my view, that an explanation offered using these styles of reasoning is, much like Putnam's peg and hole, "a full explanation,...an explanation that prescinds entirely from the stuff that genes are made from" (35): citing micro-properties is otiose. Indeed, and this will become important in a moment, the above reasoning prescinds even from the fact that the paired members are genes: if "junk DNA" is paired up in the way outlined above, it too will behave in accordance with the above abstract model.

This simple, elegant, and correct argument is surrounded by troublesome and misleading detail. Kitcher says, for instance, that though philosophers identify scientific theories with laws, there are very few laws about genes to be found in textbooks of classical genetics. One might think that "Mendel's Laws" are laws, he says, but in fact they are false: genes do not segregate independently, they do so only if they are very far apart. It is difficult to understand how to evaluate this. It is true, of course, that molecular biology will fail to reduce Mendel's second law, since the latter is false. But Mendel's second law is not a part of classical genetics. The construction of gene maps and the use of these maps in predicting trait frequencies depends on knowing that genes are arrayed on chromosomes like "beads on a string." Mendel knew nothing of chromosomes, and his law assumes that genes are not tied together. So the very practice of classical genetics depends on repudiating Mendel. Thus, it is unclear what relevance the falsity of Mendel's laws has in this context.

---

[4] Putnam took this to be an argument against the *identity* of psychological and underlying neurophysiological states. I am taking it simply as an argument that micro-level properties need not be involved in the explanation of every phenomenon.

Putting this aside, notice that the style of reasoning embodied in classical genetics is nevertheless very much like Mendel's concerning independent assortment and segregation, only more complicated because it is informed by the realization that genes do not float around independently but are rather tied together on chromosomes. Both Mendelian genetics and classical genetics are addressed to what Kitcher calls "pedigree problems," that is, problems concerning observed trait distributions, given the constraints of what is known about the location of genes (12–14). If Mendel's axioms are the kinds of propositions that would count as laws but for their falsity, why should the statements about genes and chromosomes presupposed by the classical forms of reasoning be denied this status? Certainly not because the so-called laws of classical genetics do not possess the lapidary form of Mendel's propositions: we should not expect the laws of the special sciences will have the simplicity and universality that we find in physics. That genes are particles, that genetic inheritance normally leaves Mendelian factors intact, that genes occur on chromosomes, that in sexual reproduction there is meiosis and cross-over, that some genes are operons which regulate transcription while others code for traits: all such statements should have nomic status if Mendel's "laws" could have it. Kitcher himself makes this point in a different context, when he refers to "low-level laws of chemistry," such as "DNA molecules contain adenine and thymine molecules in (almost) equal numbers" (116). He remarks in connection with such propositions: "While they are more particular than the grand equations of physics, these generalizations are scientifically significant, and are featured in numerous explanations" (116). In this light, it is hard to see why he should say that "it is hard to pick out *any* laws about genes" (5).

Kitcher suggests that molecular genetics has made it increasingly evident that there is no single kind of entity corresponding to the classical gene-concept. Thus, various statements made by classical geneticists about genes receive no univocal translation into the language of molecular genetics. Without going into the details, he is right about this.[5] Molecular genetics demonstrates that in certain contexts, the classical term "gene" has to be replaced by less inclusive terms like "structural gene," "operon," "non-coding sequence," and so on. In other contexts, for example in the discussion of independent assortment and segregation above, the term must be replaced by more inclusive terms like "DNA segment." However (as noted above) the abstract framework of classical genetics is, as Kitcher delineates it, independent of specific reference to *genes*: thus, it should work also for these new terms. Kitcher argues that classical genetics retained its identity despite correction and change in the early part of the twentieth century because through all this change it continued to investigate pedigree problems by means of the abstract model he outlines. Thus, he ought to allow that classical genetics continues to be valid in the domain of entities envisaged by molecular genetics. It is still possible to map chromosomes in the traditional way, though that method is of course losing relevance in the light of gene-sequencing and associated methods.

A lacuna in Kitcher's work on levels in biological science is that it lacks a substitute for bridge laws as classically conceived. He argues that one cannot, for every term such as *gene* that belongs to a nonphysical science (such as classical genetics), provide a statement of necessary and sufficient conditions for identifying this concept in physical terms. In any case, it is unclear what modal and epistemic status one would accord to such a statement even if it were available. As we have seen, genes do not assort the way they do because of how they are physically realized. Moreover, classical reasoning does not depend on the unity of underlying physical realizations. This leads us to an unanswered question. It has been discovered that there are functionally diverse instantiations of the old gene-concept. How exactly should this lead us to revise the old concept? We have argued that the diversity of genes has no impact on some uses of the concept in the higher-level science; at the same time, it could be that it demands changes in other uses of the same concept. Bridge laws, which have always been understood as bottom-up ontological clarifications necessary for the fully informed scientific use of higher-level concepts, are an obstacle to retaining a nonphysical concept for certain uses and abandoning it for others. Thus, bridge laws do not correctly capture the relationship between levels. But what is the appropriate alternative? Kitcher clearly wants to acknowledge and accommodate a diversity of relationships between scientific levels, but he does not offer us a systematic formal analysis that helps to adjudicate inter-level questions.[6]

---

[5] See Rafael Falk, "What Is a Gene?" *Studies in the History and Philosophy of Science*, XVII (1986): 133–73; and now Evelyn Fox Keller, *The Century of the Gene* (Cambridge: Harvard, 2000).

Species. Kitcher advances two main theses concerning species. First, that they are *sets* of specimens, and not individuals. Here he doubly contradicts Michael Ghiselin and Hull,[7] who claim that species are individuals, and not sets. Second, that there is no unique or privileged way of saying which sets count as species, or, as recent philosophers of biology would put it, that there is no unique definition of the species-concept. He calls this position *pluralistic realism.* I shall argue that Kitcher is wrong to deny that species are individuals; my argument will undercut one aspect of his pluralism.

There is a strong reason for making species *kinds* or *universals.* It is that they are used, much as ancient biological tradition supposes, to explain properties of the specimens that belong to them. Ghiselin and Hull think that individual specimens are *parts* of species. The whole-part relation does not accommodate the downward explanatory move so easily. No doubt, this point can be finessed. But it does provide us with some motivation to stick with the traditional view. On the other hand, there is a *compelling* reason for making species individuals. It is that species are causal players in evolution. By this, I do not mean just that *specimens* (individual members of species) are causal players by virtue of belonging to the species. I mean something stronger, namely that the collection of specimens is itself a causal player. I shall defend this thesis later.

There is a simple way to accommodate both positions: you distinguish between species-populations and species-kinds. The species-population is defined as follows. First, we note that a species-population is a collection of specimens. Second, we note that the specimens in such a population are linked together in such a way as to form an encompassing unity that acts as a causal factor in evolution. This allows us to define the population. Suppose, for example, that the required linking relation is that the population be a "gene pool" within which genetic recombination can occur. Then, it is true of each organism *y* that all the organisms that inhabit the same gene pool as *y* jointly constitute the species population to which *y* belongs. Thus, you can define the human population H_P as consisting of those entities that inhabit the same gene pool as certain ostended human specimens. Having defined the human population in this way, we define the species-kind *homo sapiens* as the property that humans have in virtue of being parts of H_P.

---

[6] In Mohan Matthen and André Ariew, "Two Ways of Thinking about Fitness and Natural Selection," this JOURNAL, XCIX, 2 (February 2002): 55–83, it is proposed, specifically in connection with evolutionary theory, that lower-level entities are *realizations* of higher-level entities in the sense that digital cameras are realizations of *camera* and combination locks of *lock.* A suitably generalized account of realization would, I believe, show how higher-level concepts apply directly to physical and other lower-level entities without the need for bridge laws.

[7] Ghiselin, "A Radical Solution to the Species Problem," *Systematic Zoology,* XXIII (1974): 536–44; Hull, "Are Species Really Individuals?" *Systematic Zoology,* XXV (1976): 174–91.

This construction of species-populations meets most of what Ghiselin and Hull want to say about species as individuals. The construction of species-kinds from species-populations also goes a certain distance with Kitcher in that it makes the species-kind a "set" to which organisms belong, not a mereological sum of which they are parts. It should be noted, however, that, in practice, what one might call *linking conditions* will generally be a matter of degree. For example, Ernst Mayr's "panmixis" criterion—the condition that any organism in a population may mate with any other of the opposite sex with equal probability—is intended to be a linking condition, but, as Mayr himself recognizes, it is never fully realized, but comes only in degrees of approximation. (Indeed, assortative mating, or *sexual selection* as it is usually called, is a major driver of evolution.) Because all plausible linking-conditions are like this, species-populations will be what one might call "vague entities." This means, regrettably, that there will be the following kinds of formal embarrassment: periods of time when they are not determinately either in or out of existence, temporal gaps in existence, overlaps among and nonequivalency within species-populations, indeterminate identity conditions, and so forth. Kitcher spends nearly all of chapter 6 arguing that such problems in the species-as-individual thesis cannot adequately be "faced and overcome" (145). However, such vagaries apply equally to other such entities as corporations, societies, nations, and the like, and most likely to organisms as well. What is required, therefore, is formal not biological innovation: an adequate formalism for vague entities, not a repudiation of scalar linking conditions.

Kitcher's own response to these difficulties is "pluralism." He actually articulates two versions of this principle, one in chapter 5, the other in chapter 6. In chapter 5, he takes a species to be a group of organisms that clump together in biomorphological space, differentiated from other organisms by a wide gap. He argues that there could be several different ways of drawing such groups—for example, along historical, evolutionary, ecological, or structural grounds—and hence many different species-concepts. In chapter 6, on the other hand, he is concerned with segmenting temporally extended lineages, claiming that "there is no single, objectively right, way to segment the entire lineage into species." Notwithstanding vague objects, let us simply accept that the second of these claims is at least possibly true. What about the first? Does the existence of cross-cutting similarities among organisms show that there are cross-cutting species?

In claiming that divergent scientific agenda yield cross-cutting biological classifications, Kitcher is, I believe, attaching too much importance to the similarities that unite the members of a species. He thinks that taxonomy proceeds by first carving biological morphospace at the observed joints, and then finding underlying explanatory similarities. What he and others have failed to notice is that since species are *polymorphic*, they are constituted not by one clump in morphospace, but several.[8] Consider sexual dimorphism: the human population forms two clusters divided by an sparsely occupied area of morphospace. True, men and women resemble each other more closely than either resembles a chimpanzee. But this is not sufficient for making them a single species: the various species of mosquito resemble each other more than men and women do. Why, then, are men and women not separate species? Because their separateness and interactions constitute a part of how the human population interacts with its environment and thus evolves. Note that this is not equivalent to the claim that male-human and female-human characteristics are separately selected (though this can happen too). The claim is that the whole group has characteristics that arise from dimorphism and which in turn contribute through selection to dimorphism. This is why the causal role of the whole group is indispensable; this is why the group is properly considered an individual. The division of humans into sexes with these characteristics forms part of the array condition that influenced human evolution. Notice, by the way, how this argument undercuts the species-as-single-morphoclump conception.

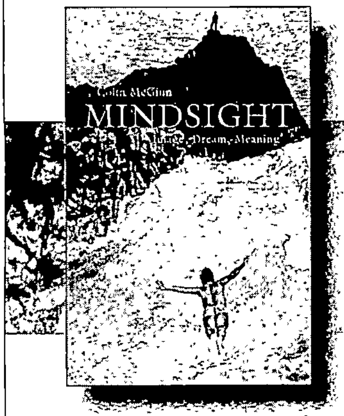
This brings us to a crucial point. Not just any morphoclump is a species: no matter how important an explanatory grouping such a clump might be, it may not be active as a single composite entity. Kitcher says: "Organisms may be divided into species according to their possession of a common 'software', and this division might cut across the distinctions drawn by attending to genetic 'hard wiring'" (123). I disagree. I am claiming that species are kinds that correspond to causally active group-individuals that have individual organisms as parts. *Only* kinds that display a population-kind duality can be species. In other words, only those kinds that are definitionally tied to a particular population are species. The collection of individuals with similar genetic software is not tied to a particular population; the condition that unites them is descriptive, not referential. Such collections are not causally active in the way that species are; they are not historically delimited in the way that species are. That certain individuals have the same software is not a linking condition that renders the group-as-a-whole a causal player, thus making it a population. This is why it does not correspond to any species. This is the point that Kitcher's species-pluralism overlooks.

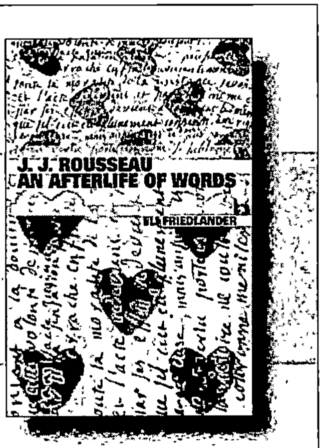
---

[8] My brief remarks on polymorphism draw on a discussion in Marc Ereshefsky and Matthen, "Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory," *Philosophy of Science* (forthcoming).

MOHAN MATTHEN

University of British Columbia





MINDSIGHT

Image, Dream, Meaning

COLIN MCGINN

Clearly, seeing an object is similar in certain respects to forming a mental image of it, but it is also different. McGinn shows what the differences are, arguing that imagination is a sui generis mental faculty. He goes on to discuss the nature of dreaming and madness, contending that these are primarily imaginative phenomena.

new in cloth

LAW'S QUANDARY

STEVEN D. SMITH

Steven Smith argues that our legal vocabulary and methods of reasoning presuppose classical ontological commitments that were explicitly articulated by thinkers from Aquinas to Coke to Blackstone, and even by Joseph Story. But these commitments are out of sync with the world view that prevails today in academic and professional thinking. So our law-talk thus degenerates into "just words"—or a kind of nonsense.

new in cloth

J. J. ROUSSEAU: AN AFTERLIFE OF WORDS

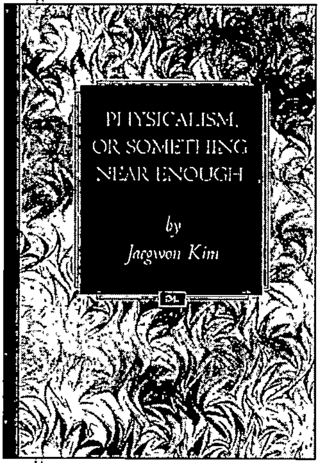
ELI FRIEDLANDER

Eli Friedlander reads Rousseau's autobiography, *Reveries of the Solitary Walker,* as philosophy. Reading this work against Descartes's *Meditations,* Friedlander shows how Rousseau's memorable transformation of experience through writing opens up the possibility of affirming even the most dejected state of being and allows the emergence of the innocence of nature out of the ruins of all social attachments.

new in cloth

WWW.HUP.HARVARD.EDU HARVARD UNIVERSITY PRESS

Notable



PHYSICALISM, OR SOMETHING NEAR ENOUGH

Jaegwon Kim

Contemporary discussions in philosophy of mind have largely been shaped by physicalism, the doctrine that all phenomena are ultimately physical. Here, Jaegwon Kim presents the most comprehensive and systematic presentation yet of his influential ideas on the mind-body problem.

Cloth $26.95 ISBN 0-691-11375-0

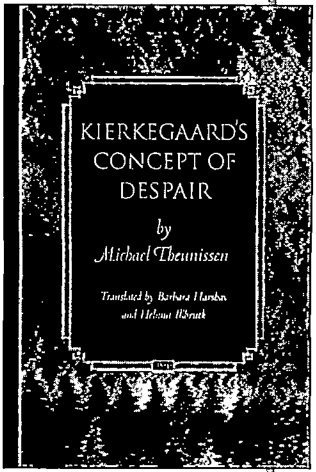
KIERKEGAARD'S CONCEPT OF DESPAIR

Michael Theunissen

Translated by Barbara Harshav and Helmut Illbruck

The literature on Kierkegaard is often content to paraphrase. By contrast, Michael Theunissen articulates one of Kierkegaard's central ideas, his theory of despair, in a detailed and comprehensible manner and confronts it with alternatives. It is also one of the few works on Kierkegaard that bridge the gap between the Continental and analytic traditions in philosophy.

Princeton Monographs in Philosophy
Cloth $35.00 ISBN 0-691-09558-2 Due May



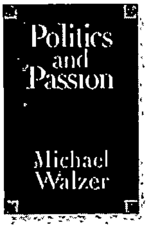
100 *Celebrating 100 Years of Excellence*
PRINCETON *University Press*
800-777-4726
Read excerpts online
www.pup.princeton.edu

Philosophy from YALE

Politics and Passion

Toward a More Egalitarian Liberalism

MICHAEL WALZER



"This **clearly and engagingly written** book shows up the thinness of much contemporary Liberal theory. In its search for the ultimate, basic principles of Liberalism, this theory neglects whole strata of actually living Liberal societies." —Charles Taylor $25.00

Also by Michael Walzer:
Arguing About War $25.00

Edmund Husserl and Eugen Fink

Beginnings and Ends in Phenomenology, 1928–1938

RONALD BRUZINA

"This book constitutes **a watershed in our understanding of the phenomenological movement,** and should become a basic reference work for all workers in the field."—Theodore Kisiel

Yale Studies in Hermeneutics $45.00

New in Paper

The Mask of Enlightenment

Nietzsche's Zarathustra

Second Edition

STANLEY ROSEN

With a new foreword by Michael Allen Gillespie

This landmark study "represents **an important contribution** to the discussion of Nietzsche as political philosopher."—Kathleen Higgins

$22.00 paperback

An Interpretation of Religion

Human Responses to the Transcendent

Second Edition

JOHN HICK

"This book strengthens Hick's position as **one of the most significant thinkers of the second half of the twentieth century. . . .** I highly recommend [it] to students of philosophy." —Chester Gillis, *Journal of Religion* (on the first edition)

$28.00 paperback

Kant's Politics

Provisional Theory for an Uncertain World

ELISABETH ELLIS

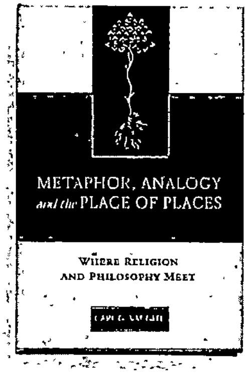
"Ellis's key contribution is her argument that takes Kant's theory of provisional right as his central political concept. **Reading Kant's political theory in any other way from now on will be difficult."** —Mika LaVaque-Manty $40.00

Hegel, Idealism, and Analytic Philosophy

TOM ROCKMORE

"Assuming (as I do) that the future of analytic philosophy requires a close attention to Hegel, I think that **Rockmore's book comes fairly close to being indispensable."** —Joseph Margolis $40.00

YALE University Press • yalebooks.com



BAYLOR UNIVERSITY PRESS
New for Spring 2005

Metaphor, Analogy and the Place of Places

Where Religion and Philosophy Meet | Provost's Series

CARL G. VAUGHT

(Ph.D. Yale) is Distinguished Professor of Philosophy at Baylor University.

ISBN 1-932792-07-4 | $29.95
6 x 9, 222 pages | Cloth

Metaphor, Analogy and the Place of Places *makes Vaught's position a major contender in contemporary philosophical theology and philosophy of religion. After reading this book, no one can claim that philosophy and religion are not integral to each other.* — **ROBERT C. NEVILLE**

baylorpress.com | 1.800.710.3217

THE JOURNAL OF PHILOSOPHY

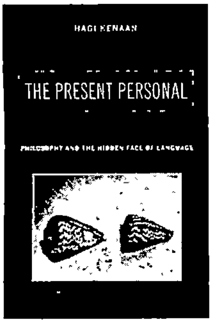
The Journal is pleased to announce to our readers and subscribers the launch of its new website, which will have the pages listed below. The address is: *www.journalofphilosophy.org* or simply *www.jphil.org.*

~•~

PURPOSE & HISTORY
CONTACTS AND ADDRESSES
BOARD OF TRUSTEES
EDITORIAL BOARD
MEMOS TO AUTHORS
SUBSCRIPTIONS & ADVERTISING
RECENT & FORTHCOMING ISSUES
ELECTRONIC ISSUES
INDEX 1904–PRESENT
RIGHTS AND PERMISSIONS
LECTURES & CONFERENCES
PHILOSOPHY LINKS

~•~

COLUMBIA UNIVERSITY PRESS



The Present Personal
Hagi Kenaan

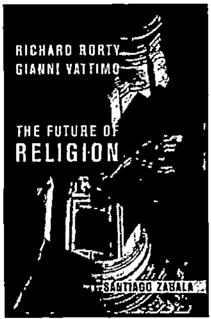
Arguing that philosophy's concern with the abstract nature of language has obscured the singular human voice, Kenaan explores the philosophical implications of recognizing and responding to the individual presence in language.

"A very impressive work by an emerging philosophical talent."—Simon Critchley, New School University

Thinking with Animals
New Perspectives on Anthropomorphism
Lorraine Daston and Gregg Mitman

This book explores the uses of anthropomorphism and its transformative effects on both humans and animals alike.

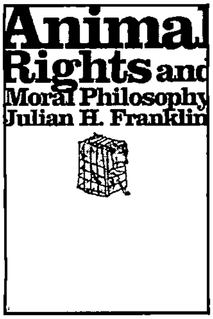
"The most original and subtle exploration of anthropomorphism to date."—Jonathan Burt, co-founder of the UK Animal Studies Group



The Future of Religion
Richard Rorty and Gianni Vattimo
Edited by Santiago Zabala
Translated by Wm. McCuaig

Two prominent philosophers explore the role of religion and belief in contemporary society.

"A delightful dialogue that challenges the beliefs of theists and atheists alike." —Jean Grondin, University of Montreal



Animal Rights and Moral Philosophy
Julian H. Franklin

This theoretically rigorous book examines all of the major arguments for animal rights proposed to date and extends the theory in new directions.

"A valuable contribution to a vital debate."—J. M. Coetzee, winner of the Nobel Prize in Literature, 2003

Kant's Philosophies of Judgement
Douglas Burnham

This book explores Kant's three philosophies of judgement in a new and creative reading of Kant's work and of the concept of judgement in general.

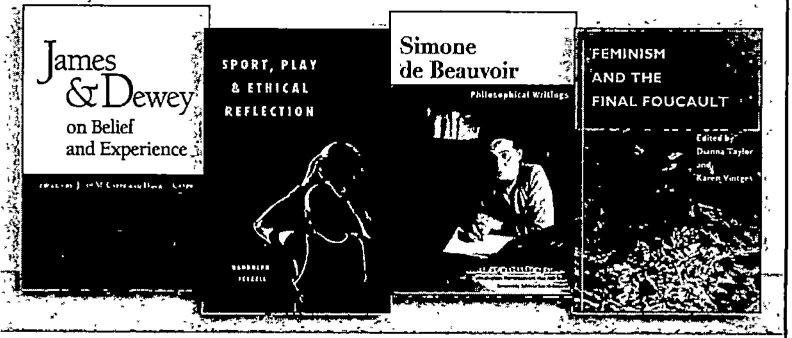
Expanded Edition
Nietzsche as Philosopher
Arthur C. Danto

The expanded edition of this influential study of Nietzsche from a leading contemporary philosopher now features a new preface and five new essays.

columbia.edu/cu/cup

UNIVERSITY OF ILLINOIS PRESS

To order, call 800-537-5487, or visit www.press.uillinois.edu


James & Dewey on Belief and Experience
SPORT, PLAY & ETHICAL REFLECTION
Simone de Beauvoir
Philosophical Writings
FEMINISM AND THE FINAL FOUCAULT
Edited by Dianna Taylor and Karen Vintges

James and Dewey on Belief and Experience

Edited by JOHN M. CAPPS
and DONALD CAPPS

Juxtaposes the key writings of these two fathers of Pragmatism, America's unique contribution to world philosophy. An editors' introduction provides biographical information about both men, an overview of their respective philosophical orientations, a discussion of the editorial process, and a brief commentary on each of the selections.

Cloth, $45.00; Paper, $20.00

Sport, Play, and Ethical Reflection

RANDOLPH FEEZELL

Feezell argues that sports participation is best described as a form of human play, arguing that sportsmanship should be seen as a kind of Aristotelian mean between the extremes of over- and under-investment in sport.

Cloth, $30.00

Philosophical Writings

SIMONE de BEAUVOIR

Edited by Margaret A. Simons with Marybeth Timmermann and Mary Beth Mader

Foreword by Sylvie Le Bon de Beauvoir

The first complete, scholarly edition of Beauvoir's essays in English translation highlights continuities in the development of her thought. Covering the first twenty-three years of her work, this book includes some essays only recently discovered.

Inaugural volume in The Beauvoir Series

Cloth, $40.00

Feminism and the Final Foucault

Edited by DIANNA TAYLOR
and KAREN VINTGES

This volume illustrates the ways in which Foucault's later thinking on ethics as "care of the self" can reintroduce a number of issues and themes that feminists jettisoned in the wake of postmodernism, including consciousness raising, feminist therapy, the subject woman, identity politics, and feminist agency.

Cloth, $50.00; Paper, $20.00

Embodied Care

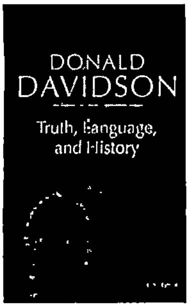
Jane Addams, Maurice Merleau-Ponty, and Feminist Ethics

MAURICE HAMINGTON

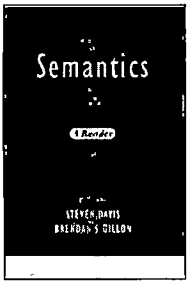
The first work to argue for the body's centrality to care ethics, doing so by analyzing our corporeality at the phenomenological level. Hamington takes the reader on a journey from modern care ethics to Merleau-Ponty's philosophy of the body and then to Jane Addams's social activism and philosophy.

Cloth, $30.00

New from
Oxford

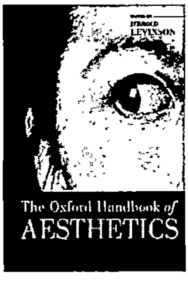


Truth, Language, and History
Philosophical Essays
Volume 5
DONALD DAVIDSON
Truth, Language, and History is the much-anticipated final volume of Donald Davidson's philosophical writings. In four groups of essays, Davidson continues to explore the themes that occupied him for more than fifty years: the relations between language and the world; speaker intention and linguistic meaning; language and mind; mind and body; mind and world; mind and other minds.
April 2005 paper $27 50 cloth $85.00



Semantics
A Reader
Edited by STEVEN DAVIS AND BRENDAN S. GILLON
Semantics: A Reader is a collection of published articles in the area of semantics, and is designed as a textbook for use in upper level philosophy and linguistics courses and as a reference for scholars of linguistics who want the classic articles in their field in one convenient place.
2004 paper $45.00 cloth $85 00

Joint Attention: Communication and Other Minds
Issues in Philosophy and Psychology
Edited by NAOMI EILAN, CHRISTOPH HOERL, JOHANNES ROESSLER, and TERESA MCCORMACK
This volume brings together, for the first time, psychological and philosophical perspectives on the nature and significance of joint attention. Original contributions by leading researchers in both disciplines explore the idea that joint attention has a key foundational role to play in the emergence of communicative abilities, psychological understanding, and, possibly, in the very capacity for objective thought.
April 2005 paper $39 95 cloth $99.00



New in Paperback!
The Oxford Handbook of Aesthetics
Edited by JERROLD LEVINSON
"Levinson has achieved his intention to provide a collection from which both the professional philosopher and the enthusiastic non-professional can derive instruction and pleasure. In doing so, he has also succeeded in filling a sizeable gap in a burgeoning book market."—*British Journal of Aesthetics*
paper April 2005 $45.0C

Art and Intention
A Philosophical Study
PAISLEY LIVINGSTON
Do the artist's intentions have anything to do with the making and appreciation of works of art? In *Art and Intention* Paisley Livingston develops a broad and balanced perspective on perennial disputes between intentionalists and anti-intentionalists in philosophical aesthetics and critical theory.
April 2005 $65 00 JKC 12505

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

CAMBRIDGE

NEW AND NOTEWORTHY

The Logical Foundations of Bradley's Metaphysics
Judgment, Inference, and Truth
James W. Allard
$75.00: Hardback 0-521-83405-8: 262pp

The Rights of Others
Aliens, Residents and Citizens
Seyla Benhabib
The Seeley Lectures 5
$65.00: Hardback 0-521-83134-2: 264pp
$23.99. Paperback· 0-521-53860-2

Plato's Introduction of Forms
R.M. Dancy
$75.00: Hardback: 0-521-83801-0: 360pp

The Creativity of God
World, Eucharist, Reason
Oliver Davies
Cambridge Studies in Christian Doctrine 12
$75 00: Hardback· 0-521-83117-2: 222pp
$27.99: Paperback: 0-521-53845-9

Justice and Democracy
Edited by Keith Dowding, Robert E. Goodin, and Carole Pateman
$70.00: Hardback 0-521-83695-6: 242pp
$25.99: Paperback: 0-521-54543-9

Myths of the Underworld Journey
Plato, Aristophanes, and the "Orphic" Gold Tablets
Radcliffe G. Edmonds III
$75.00. Hardback· 0-521-83434-1 288pp

Closing the Books
Transitional Justice in Historical Perspective
Jon Elster
$70.00: Hardback· 0-521-83969-6: 310pp
$24.99: Paperback. 0-521-54854-3

Machiavelli and Empire
Mikael Hornqvist
Ideas in Context 71
$75.00: Hardback: 0-521-83945-9. 320pp

Crimes Against Humanity
A Normative Account
Larry May
Cambridge Studies in Philosophy and Law
$70.00: Hardback: 0-521-84079-1: 326pp
$25 99: Paperback: 0-521-60051-0

Problems and Methods in the Study of Politics
Edited by Ian Shapiro, Rogers M. Smith, and Tarek E. Masoud
$75.00: Hardback: 0-521-83174-1: 432pp
$29 99· Paperback: 0-521-53943-9

Second Edition!
Chomsky
Ideas and Ideals
Neil Smith
$70 00. Hardback: 0-521-83788-X: 296pp
$27.99. Paperback: 0-521-54688-5

Personal Autonomy
New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy
Edited by James Stacey Taylor
$70 00: Hardback: 0-521-83796-0: 360pp

The Phenomenology of Painting
Nigel Wentworth
$80.00. Hardback: 0-521-81999-7: 296pp

www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

University
04/01

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 5
MAY 2005

page
217 *Toward a Projectivist Account of Color* Edward Wilson Averill
235 *Linear Perspective as a Realist Constraint* Anthony A. Derksen
BOOK REVIEWS
259 Alva Noë: *Action in Perception* Ned Block

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721. You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc.

ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 5, MAY 2005

TOWARD A PROJECTIVIST ACCOUNT OF COLOR*

In this paper, seeing is assumed to be a matter of visual representation. Thus, seeing something as having a color is a matter of visually representing it as having a color. Here 'color', and the more determinate terms 'red', 'orange', 'brown', 'white', and so on, are used to refer to properties we visually attribute to things. The nature of these properties is obscure. It is not at all clear how they fit into a naturalistic account of things, or even what sorts of objects, if any, are colored. In any case, a theory of color, as understood here, seeks to clarify the nature of color by setting out an ontology of the color properties and explaining how they are related to vision. Such a theory is not, or at least not directly, a theory about our ordinary use of color terms.

Color theories are conveniently divided into those that hold that colors are independent of visual systems, *vision independent theories*, and those that deny this claim, *vision dependent theories*. Since independence is a negative idea, it is best to start with dependence. By definition, *A* is dependent on *B* if and only if some variation in *B* is sufficient to bring about a variation in *A*. So a vision independent theory holds that no variations in our visual mechanisms would be sufficient to bring about a variation in the colors of physical objects.[1] The most commonly defended form of vision independent theories, *color objectivism*, adds a perceptual assumption: our perceptual representations of objects as col-

---

* I would like to thank the following people for the help they gave me in connection with this essay: Jonathan Cohen, Larry Hardin, David Hilbert, Mohan Matthen, and Robert Rupert.

[1] A vision independent theory does not preclude there being a necessary connection between the colors of things and vision. For example, some philosophers have held that, necessarily, *x* is red if and only if *x* would look red under standard conditions. If 'standard conditions' rigidly refers to some set of conditions as they presently occur, colors are independent of visual systems because a change in visual systems would not change the colors of things. Of course if 'standard conditions' is taken nonrigidly so as to allow that the colors of things can change when standard perceivers change, then the position holds that colors are vision dependent.

0022-362X/05/0205/217–34

ored—as red, or blue, or greenish yellow, and so on—is, in many cases, not only veridical but is caused in part, and in the right way, by the color of the object so that the perceiver is justified in believing that the color her visual system represents the object as having is the color of the object. For example, we are justified in believing that ripe tomatoes are red.

Color variation arguments against objectivism begin with the claim that it is possible for the same object to appear quite different in color to different perceivers. The trick is to show that this fact undermines objectivism. In section I, a color variation argument is used to bring out limitations in the objectivist account of color developed by Alex Byrne and David Hilbert.[2] Some objectivist accounts accommodate color variation arguments by allowing that an object can instantiate many different color properties at one time, which makes it possible for two observers with different visual systems to see different color properties of the same object at the same time.[3] Relational and projectivist accounts are responses to color variation arguments that reject at least one of the objectivist assumptions. Relational accounts are vision dependent theories.[4] Projectivist accounts reject the perceptual assumption. In section II, I shall argue for a projectivist account of color.

I. THE OBJECTIVISM OF BYRNE AND HILBERT

To get at the representational content of our color experiences, Byrne and Hilbert make the plausible assumption that colors are physical properties that underlie the causal processes of the perception of color, and set themselves the task of identifying these properties (CRCS 8). They begin their study by restricting it to the perception of matte surfaces, and say "We can narrow the field further by noting that the color vision of human beings and many other organisms exhibits approximate color constancy" (CRCS 9). This is the first use of the term 'color vision' in the development of their theory. They are "narrowing of the field" in two ways: they are only considering matte surfaces, and they are only considering a certain type of vision, color vision, which excludes night vision. Of course there is nothing in the original problem that requires this narrowing.

---

[2] "Color Realism and Color Science," *Behavioral and Brain Sciences*, XXVI, 1 (2003): 3–63 (hereafter CRCS). There are many versions of objectivism. These include D.M. Armstrong, *A Materialist Theory of Mind* (New York: Humanities Press, 1968); J.J.C. Smart, "On Some Criticism of a Physicalist Theory of Colors," in C.Y. Cheng, ed., *Philosophical Aspects of the Mind-Body Problem* (Honolulu: Hawaii UP, 1975), pp. 54–63, reprinted in Byrne and Hilbert, eds., *Readings on Color: The Philosophy of Color*, Volume 1 (Cambridge: MIT, 1997), pp. 1–10; Edward Wilson Averill, "Color and the Anthropocentric Problem," this JOURNAL, LXXXII, 6 (June 1985): 281–304, reprinted in Byrne and Hilbert, eds., pp. 11–32; Hilbert, *Color and Color Perception: A Study in Anthropocentric Realism* (Stanford: CSLI, 1987), and "What Is Color Vision?" *Philosophical Studies*, LXVIII (1992): 351–70; David K. Lewis, "Naming the Colors," *Australasian Journal of Philosophy*, LXXV (1997): 325–42; Frank Jackson, *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (New York: Oxford, 1998); Michael Tye, *Ten Problems of Consciousness; A Representational Theory of the Phenomenal Mind* (Cambridge: MIT, 1995), and *Consciousness, Color, and Content* (Cambridge: MIT, 2000).

[3] For an example, see Mohan Matthen, "The Disunity of Color," *The Philosophical Review*, CVIII, 1 (1999): 47–84.

[4] For examples of relationalist theories, see Averill, "The Relational Nature of Color," *The Philosophical Review*, CI, 3 (1992): 551–88; and Jonathan Cohen, "Color Properties and Color Ascriptions: A Relationalist Manifesto," forthcoming in *The Philosophical Review*.

Humans tend to represent visually the matte surface of an object as having the same color under many different lighting conditions. These color constancy facts lead Byrne and Hilbert to hold that, for humans, representing the surface of an object as having a color is representing the object as having a surface feature that is independent of illumination. There are only two plausible candidates for such a surface feature: the spectral reflectance of the surface and the categorical basis in the surface for its spectral reflectance. For the purposes of this essay, it does not matter which is taken to be the color of a surface. Therefore, I follow the Byrne and Hilbert choice of the former feature.[5] Hence, on their theory, a human visual representation of a matte surface as colored is a representation of the surface as having a spectral reflectance of a certain type (the type being different for different colors), even if the human does not realize this.[6] A color—red, orange, brown, and so on—is a type of spectral reflectance. (Byrne and Hilbert generalize the above theory from objects with opaque matte surfaces to include transparent and translucent objects, as well as light sources. To simplify matters, I shall ignore this generalization and argue that their theory will not work for matte surfaces. They never generalize their theory to include night vision.)

To fill out this theory, the spectral reflectance types of the different hues—red, blue, green, and so on—and their shades need to be identified. Let us start with hues, taking yellow as an example. Since there are many shades of yellow, yellowness is a set of spectral reflectances. But how are the members of this set determined? On the theory of Byrne and Hilbert, the content of our experiences of yellow determines the set of spectral reflectances. If a naturalized semantics were available, it would identify the spectral reflectances that are, and are not, members of the yellow type. But no such semantics is available and hence there is no way of determining the set of reflectances that constitute yellowness (CRCS 8, 57).

Shades of yellow are also sets of spectral reflectances—subsets of yellowness—due to metamers. Two objects that have different spectral reflectance curves are metamers for observer *S* under lighting conditions *C* if they look exactly the same in color to *S* under *C*. Many objects are metamers to all human observers in sunlight. (And many objects are metamers for some observers in sunlight but not for others.) Thus, two different objects, which look to be the same shade of yellow in sunlight, can be reflecting light at very different wavelengths. So a specific shade of yellow will be identical to a set of spectral reflectance curves; similarly for other colors. About any specific yellow surface we can ask "What is the set of spectral reflectances that constitute this shade of yellow?" Byrne and Hilbert cannot tell us how to find out because this set is determined by the content of our (veridical) experience of this shade, and without a naturalized semantics there is no way to determine this content.

---

[5] Jackson holds that it would be better to take the latter as the color of a surface, because spectral reflectances cannot do causal work. The color variation objection raised here to Byrne and Hilbert's theory also applies to Jackson's theory.

[6] The way in which the color of an object, its spectral reflectance, is represented in color vision (its mode of presentation) does not reveal to the viewer the nature of the color.

As I understand Byrne and Hilbert, they assume that the colors of many objects are known prior to color theory—we know that ripe tomatoes are red and that gold of the impure kind used in jewelry is yellow. To deny that we have such knowledge would lead to a skepticism that they reject (CRCS 57). So I read Byrne and Hilbert as holding that the yellowness set includes as a subset the spectral reflectances of most of the objects we commonly identify (or would identify) as yellow. I shall now argue that, given the Byrne and Hilbert theory, we do not know that ripe tomatoes are red and that gold jewelry is yellow.

To get at the problem, I need to say a bit about how we perceive colors and how humans with a slightly different perceptual mechanism would perceive colors. Among the receptor cells in the human retina are three types of cone cells, known as the S, M, and L cells. (The retina also contains rod cells for night vision.) Each of the cone cells has a pigmented tip that absorbs light, making the cell sensitive to the light absorbed. The S, M, and L cells differ in the kinds of pigments they are tipped with and thus the kinds of light they absorb. The S cells are most sensitive to light around 440nm, the M cell to light around 530nm, and the L cells to light around 560nm.

Now, suppose some isolated group of humans—space travelers on a distant planet with an earth-like environment—were to evolve (over thousands of years) so that the pigments of the M cells in their retinas had a slightly different absorption curve (the modified M cells are less sensitive in the 540 to 600nm range). And suppose that this caused the following condition: when space travelers see gold under (normal conditions), the mode of presentation of the color property they represent visually the gold as having is the same as the mode of presentation of the color property contemporary humans visually represent ripe tomatoes as having, when they see ripe tomatoes (under normal conditions).[7]

---

[7] To get a better idea of how this might work, consider some formulas relating the stimulation of L, M, and S cone cells to the colors we see. Very roughly $(L\text{-}M) > 0$

(A suggestive, but question begging, way of putting this point: gold looks red to the space travelers.) There are two ways of interpreting the space traveler scenario. (A) The perceptions of gold are veridical for both contemporary humans and space travelers, because they use different systems of classification. The content of the space traveler's experience of gold attributes red_{st} to gold and the content of contemporary humans's experience of ripe tomatoes attributes red_{ch}. The reflectance types that space travelers visually represent objects as having are not colors, although in their visual systems reflectance types function much like colors function in our visual systems. Now, it seems that whether the space travelers see things as colored or not depends on a minor verbal difference on what counts as a color. (B) Both contemporary humans and space travelers use the same system of classification, but differ on how to classify gold making at least one of the parties wrong about the color of gold. There are two versions of (B). (B-1) One party is right, presumably us, and the other is wrong. (B-2) Both parties are wrong—the claim of projectivism.[8] Byrne and Hilbert could adopt either (A) or (B-1). Their discussion of the perceptions of other animals suggest that the (A) response is available to them. (In different ways both Matthen (*op. cit.*) and Cohen (*op. cit.*) take alternative (A).) However, (B-1) is more like the Byrne and Hilbert response to a similar problem that arises with shades (see below). In this case, the Byrne and Hilbert theory can be charged with a second problem:

---

codes for red, (L-M) < 0 codes for green and (L-M) = 0 codes for brain gray; and [(L+M)-S] > 0 codes for yellow, [(L+M)-S] < 0 codes for blue, and [(L+M)-S] = 0 codes for brain gray (see Stephen E. Palmer, *Vision Science: Photons to Phenomenology* (Cambridge: MIT, 1999), pp. 113–14). The modified M cells are less sensitive in the 540 to 600nm range. Consider an object that only reflects light at 545nm. At 545nm the S cells do not fire. At 545nm (L-M) before the change is zero, because the L and M sensitivity curves cross at 545nm. (L-M) after the change is positive. Also: (L+M) before the change is greater than (L+M) after the change. *Before the change*: at 545nm L=M and there is no coding for either red or green. Since S is zero, (L+M)-S is positive coding for yellow. The object looks yellow. *After the change*: at 545nm L>M, which codes for red. (L+M) is less than before the change making [(L+M)-S] code less strongly for yellow. Now the object looks yellowish red. (This example is illustrative only. In a more exact account the coding formulas would include weights for the L and M cone inputs.)

The reflectance curve for gold is more complex than this imagined object. Still (impure) gold now looks yellow, but after the change it would look red. (For other versions of this point see Averill, "The Relational Nature of Color," and Matthen.)

[8] Suppose further that the earth humans have died off. One might argue (much as Averill does in "The Relational Nature of Color") that this possibility shows that a change in the visual system of humans could change the color of gold from yellow to red. This would be a direct challenge to the assumption that colors are independent of visual systems. However, the argument as it stands is invalid because it does not show that the change in the human perceptual system was by itself sufficient to change the color of gold, since the possibility described here rests on other changes—space travel, a die off of humans. For this reason the objections raised here are more indirect.

not only do they not identify the sets of reflectances that constitute the hues (a nature of colors problem), they also do not account for why the perceptions of the colors by contemporary humans is more accurate than that of the space travelers' perceptions (a veracity of perception problem). Of course, Byrne and Hilbert are free to speculate that a naturalistic semantics of perception would show that our perceptions of the world are more accurate than those of the space travelers. However, I see no reason to believe this.

Byrne and Hilbert claim that the possibility of such space travelers is a skeptical possibility, which raises epistemological problems that a theory of color is not obliged to address (CRCS 57). But, given (A), the possibility of such space travelers does not by itself raise a skeptical point, because it does not show that our belief that gold is yellow is false. It rather shows that there are alternative ways of classifying the colors of objects. So I conclude that the Byrne and Hilbert charge that the space travelers possibility is a skeptical possibility is false. Given (B-1), of course, a skeptical conclusion looms since there seems to be no way of showing that we know that gold is yellow and not red.

So far, I have argued that Byrne and Hilbert have two hue problems: a nature-of-hues-problem and a veracity-of-perception problem. Their theory also has similar shade problems. The shade of a specific yellow color chip, call it 'Y_{18}', is a set of spectral reflectances. Among human observers who pass all the standard tests for color blindness, there are differences in the way they match shades. Thus, some color chip, call it 'Y_{22}', will appear to be exactly the same shade of yellow as chip Y_{18} to some people, but not to others, even though Y_{18} and Y_{22} have different spectral reflectance curves: chips Y_{18} and Y_{22} are metamers under sunlight for some observers, but not for others. Are Y_{18} and Y_{22} the same shade of yellow or are they different shades of yellow? Byrne and Hilbert have no way of answering this question and thus no way of identifying the set of spectral reflectance curves that constitute a specific shade of yellow. And, like the (B-1) interpretation of the space traveler argument, the problem is that there is no way of showing that, among normal human observers, some more accurately identify shades of yellowness (or any other hue) than others.[9]

Consider how this problem arises with unique green (a green that has no blue or yellow in it). As Byrne and Hilbert point out, when normal observers are asked to identify spectral lights that are unique green they pick lights that vary from 490 to 520nm. Since 15nm more than unique green looks bluish green and 15nm less looks yellowish green, this is a wide variation. So Byrne and Hilbert conclude that they do not know which color chips are unique green. There is such a shade of green according to their theory, but we cannot identify it. More exactly, given the reflectance physicalism of Byrne and Hilbert, there is some light reflecting property (or disjunction of properties) *x* such that, necessarily, unique green = *x*. However, the theory of Byrne and Hilbert does not have the resources to fill in the *x* or to show how the *x* could, even in theory, be identified (a nature of shades problem). Furthermore, their theory provides no method of determining which perceivers accurately perceive objects to be unique green (a veracity of perception problem).

---

[9] Suppose one took the view that if anyone can distinguish between two color chips then the chips are different in color. This allows there to be super observers of color, and ultimately for machines to distinguish colors. Indeed, the upshot is that every difference in reflectance is a different color. (See Averill, "Color and the Anthropocentric Problem" and Hilbert, *Color and Color Perception*.) Whatever the merits of such an account of color, it cannot, without supplementation, identify the hues and their shades. Such a theory has no contact with the color solid.

Of course, these two problems are related. Although being yellow, or a particular shade of yellow, is a vision independent property of some objects, Byrne and Hilbert cannot specify a vision independent criteria for being yellow or being a particular shade of yellow. If we had such a criterion we could, presumably, use it to determine which perceptions of objects as yellow are veridical. Without the criterion we have no way of testing the accuracy of color perceptions.

II. TOWARD A PROJECTIVIST ACCOUNT OF COLOR

Projectivism holds that colors (somehow) enter the visual system, and because of the way they enter many of the objects we see, which are not colored, look colored. A major problem for any projectivist account is to spell out how colors enter the visual system. There are two approaches. Literal projectivism holds that colors enter the visual system by being instantiated in some part of it. For example, Paul Boghossian and David Velleman hold that "...colors are qualitative properties of visual experiences that are mistakenly projected onto material objects."[10] Clearly such a claim amounts to a suggestion waiting for the nature of visual experiences to be spelled out. The figurative projectionists (such as Averill in 1992 (*op. cit.*)), have maintained that color properties are not instantiated in anything, although vision does represent objects in terms of them; that is, colors only enter the world as represented properties.[11] The problem for the figurative projectionist is to explain how it is possible for vision to represent objects as having colors if colors are not instantiated anywhere.

In what follows, a form of literal projectivism will be sketched out in terms of some of the advancements visual scientists have made in understanding visual representation. On the one hand, I think there is a lot of substance in this sketch. On the other hand, it is very clear that the way visual scientists understand visual representation is seriously incomplete. What they have to offer is more like a sketch of how visual representation might go, rather than a real theory. Consequently the projectivism developed here is incomplete in ways I shall point out.

---

[10] "Colour as a Secondary Quality," *Mind*, XCVIII (1989): 81–103, p. 103, reprinted in Byrne and Hilbert, eds., pp. 81–103.

[11] The terms 'literal projectivism' and 'figurative projectivism' come from Sydney Shoemaker, "Phenomenal Character," *Noûs*, XXVIII (1994): 21–38, reprinted in Byrne and Hilbert, eds., pp. 227–45.

One more preliminary note: part of the projectivist project is to give an account of looks red. Once this has been accomplished a linguistic rule can be adopted of this form: *x* may be described as red if and only if *x* looks (or would look) red for such-and-such observers under such-and-such conditions. The such-and-such blanks may be filled in more or less specifically to form more or less clearly defined sets of observers (normal observers, say) and conditions (standard lighting, say). Thus, the projectivist gets linguistically constructed color properties that are attributed to objects in thought and language. I have been using 'red' to refer to a property we visually attribute to objects. In what follows, I shall use '*red' when a linguistic rule of the above sort is implicitly or explicitly used to specify a property we attribute to objects in thought and language, although the appropriate account of looks red upon which this usage depends will not emerge until the end of the essay. Similar remarks hold for other colors.

II.1. Cell Firings and Colors. Vision scientists believe that the firing rates of opponent cells are correlated with color experiences. For example, R+G- opponent cells are correlated with red experiences when firing above their base firing rate, and are correlated with green experiences when firing below their base firing rate. The reverse is true for G+R- opponent cells. Similarly the firings of B+Y- and Y+B- cells are correlated with blue and yellow experiences, and the W+B- cells and the B+W- cells are correlated with experiences of black and white. The first question to ask is how should we understand the basis for these correlations.

To get a handle on this question, I make three assumptions about properties, which appear plausible although each has been disputed. First: investigations into the ways in which things causally interact with one another give us direct knowledge about the dispositional properties of objects. If these investigations give us further knowledge of the properties of objects, that knowledge is inferred using this dispositional knowledge. For example, an object with inertial mass has a disposition to accelerate, in accordance with Newton's second law, when a force is exerted upon it. Second: every disposition has a causal basis composed of one or more properties, each of which is either dispositional or nondispositional. (The causal basis of the fragility of a glass would be its molecular properties. Supporters of bare dispositions hold that some dispositions have no causal basis.[12]) Third: there is no infinite series of dispositional properties such that each is followed by one of its causal base properties. So some base properties are nondispositional. Although such properties play a very important role in causal interactions, they are not directly revealed in the causal interactions of objects. Such nondispositional base properties might explain the correlations noted above. Colors are identified by their representational role in visual experience (that is, it is a contingent fact about colors that we visually represent objects as having them), and not by their role in causal interactions and thus color properties might be base properties of opponent cell firings.[13]

The hypothesis I want to run with is that opponent cell firings have nondispositional color properties that are the basis for those firings. Call these *c-colors* to bring attention to the fact that these are color properties that are instantiated in opponent cells. When a R+G- cell fires above its base firing rate (or a G+R- cell fires below its base firing rate) the cell is c-red, and this redness is what causes the cell to fire above (or below for G+R- cells) its base firing rate; similarly, when the cell fires below (or above for G+R- cells) its base firing rate it is c-green, and its greenness causes the cell to fire below (or above for G+R- cells) its base firing rate. C-red and c-green are not classes of colors, but rather specific shades of red and green respectively. C-red is a pure unitary determinate red color that is not bluish or yellowish or blackish or whitish (it is unique red); similarly for c-green. The firings of Y+B- and B+Y- opponent cells are c-yellow and c-blue; the firings of W+B- and B+W- cells are c-white and c-black. All opponent cells are c-gray when firing at their base rates.[14]

C-colors are properties of firing opponent cells, and are some of the many colors that objects are visually represented as having. Red, unlike c-red, is a determinable that has many shades, including c-red. It is assumed here that all colors have a sensuous nature, which goes undetected by physics because this sensuousness is nondispositional. From the sensuousness of c-colors, it does not follow that opponent cell firings are in some way conscious. Rather, I assume, the sensuousness of c-colors is a metaphysically necessary characteristic of c-colors that they have independently of visual experience or consciousness. (So this is not panpsychism.)

---

[12] For a defense of this position, see Shoemaker, "Causal and Metaphysical Necessity," *Pacific Philosophical Quarterly*, LXXIX (1998): 59–77.

[13] Here is a sketch of a similar argument developed by Bertrand Russell, *The Analysis of Matter* (New York: Routledge, 2001, originally published in 1927). Physics posits objects (electrical charges, fields, massive objects) and sets out relations, including causal relations, between them. But physics does not tell us about the intrinsic nature of the objects it posits. For example, we learn that, for any point in space, the divergence of the electrical field is directly proportional to the charge density at that point. This tells us how charge density effects the electrical field but does not tell us anything about the intrinsic nature of either charges or electrical fields. Thus, for all we know, these objects could have color properties. For a discussion of Russell's point and references to other discussions of it, see David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford, 1995), and "Consciousness and Its Place in Nature," in Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings* (New York: Oxford, 2000), pp. 247–72.

[14] Although colors are the basis for opponent cell firings, perhaps colors are also the basis for other circuit firings, including nonbiological cases. This theory allows for vision independence with respect to colors—other than those that occur in opponent cell firings—but it is not objectivist.

To get at the role of colors in seeing, an account of sight is needed. Following the lead of many visual scientists, it is assumed that for an animal to see—that is, for an animal to represent visually objects in its environment—its visual system must complete two sorts of tasks: the building of a model and a classification (commonly called "categorizations" by vision scientists) of the objects modeled. I begin with visual modeling. (Seeing *x* is not the same thing as being visually conscious of *x*—see section II.3d.)

II.2. Visual Models and Colors. A visual model is a one-to-one mapping from a set of neurologically defined elements (the domain) onto a set of geometrically defined elements in the external world (the range). The elements of the domain are groups of spatially continuous nerve cells. Each such group has a size, a shape, and a location in the brain. Even the smallest of these elements is assumed to be composed of many nerve cells (only some of which are opponent cells), because classification does not seem to be so fine grained that it deals with individual cell firings. The individual cells that make up an element have firing rates and thus c-colors for the opponent cells. (Opponent cells occur throughout the visual system, including the retina and the visual cortex.) The elements of the range are continuous regions of space that have shape, size, and location relative to one another.

Domain elements are colored—a color they derive from the pattern of their opponent cell firings. If all the opponent cells of a domain element are c-green then the element is green. If an element in the domain has a large percentage of c-red cells, a smaller percentage of c-yellow cells, and both types are evenly distributed throughout the element, then the domain element is reddish orange. Central to this account of the color of domain elements is the notion of color density—in this case the c-red cells are more dense than the c-yellow cells. Domain elements are distinguished from one another, in part, by the patterns of their opponent cell firings—and thus by their colors. Hence, the boundaries between domain elements are marked by changes in color, much like changes in *colors mark the boundaries of the states in a map of the U.S. (I will say more on color density below.)

A visual map is no more than a one-to-one correspondence between elements that are unclassified and so unclassified by either geometric or color features.[15] If an element in the domain is square or blue, these features of the element are not part of what is mapped onto the range, because the map is simply a correspondence between elements. In the maps generated early on in visual processing the elements of the domain may be blobs and bars, or they may be spatially extended patterns of sinusoidal gratings. Given the theory of this paper, part of the identity of a blob or a bar as an element is due to a pattern of opponent cell firings—and thus its c-colors—that distinguish it from other elements. We do not see things in the range of such lower-level visual models—for example, we do not see things as blobs and bars in our environment—because visual classification does not take place with respect to them. In the three-dimensional (3-D) models, at the final stages of visual modeling, the elements in the domain are distinguished by their shapes, sizes, locations in the brain and their colors. It is with respect to these models that visual classification takes place.

One weakness of the hypothesis that vision depends on a mapping is that there is no good account of exactly what counts as an element in either the domain or the range.

II.3. Visual Classification. Visual classification is poorly understood. What follows is a rough description of what seems to happen. Let x be an element in the range of a visual mapping, and X the corresponding element in the domain. Classification comes about when the visual system has a dispositional reaction to X of the form: the element corresponding to X, namely x, is one of *that* kind. The kind is determined by the nature of the disposition involved. As mentioned earlier, visual representation enters at the level of classification; that is, in classifications the visual system represents x as being a member of a kind and the subject sees x as a thing of that kind. (So all seeing is seeing as.) The visual system can classify the elements in the range in many ways; by geometric properties (being a cube), functional roles (being a chair), natural kinds (being water), and colors (being red). In the duck-rabbit case the same model results in two different classifications of the range element. Projection occurs when a feature F of a domain element (or a relation R among domain elements) triggers a classification of the corresponding range element (or elements) as being an *F* (or an *R*). For example, if *X* is a red (or square) domain element it may cause a classifying disposition to classify the corresponding range element *x* as red (or square) and thus the viewer sees *x* as red (or square). Of course, it is basic to the idea of modeling that the model and the modeled share some of the same features. Notice that projection is a primitive form of seeing—we can see an animal as a fox, but this is not a matter of projection. (In the case of after-images there is an element in the domain that does not correspond to an element in the range. Like Descartes's dream, this is an illusion of visual representation.)

---

[15] But surely, it will be said, the model does some classification. At the very least a visual model is a discriminator. Since elements in the domain correspond to elements in the range, if X and Y are elements in the domain then the corresponding x and y elements in the range are represented by the model as distinct elements. This is not true. It does follow from a description of one-to-one modeling that if X and Y are distinct elements in the domain of the model then x and y are distinct elements of the range. But discrimination is still not something a visual model does, because the model does not have a mechanism for conferring the "many-one" feature that is necessary for classification. For all couples, the relation x-is-different-from-y holds, but this similarity among couples is not something that the model as a model can confer on a couple or represent the couple as having. What a model does needs to be distinguished from what a description of a model implies.

Monadic classification can be extended to dyadic classification. Let *x* and *y* be elements in the range of a visual mapping, and *X* and *Y* the corresponding elements in the domain. Again, we suppose that classification comes about when the visual system has a dispositional reaction to *X* and *Y* that attributes to *x* and *y* a relation having an indexical character—there is a relation between *x* and *y* of *that* kind, where the kind is determined by the disposition involved. That is, the visual system is representing *x* as being related to *y* in a certain kind of way—the subject sees a relation of such-and-such a kind holding between *x* and *y*. For example, if *X* is next to *Y* in the brain, *x* and *y* may be classified, and thus seen, as being next to each other—another example of a projection.

A major weakness of this theory is that there is no good account of how classifying dispositions come about. On a causal theory (very roughly), if tokens (elements) of type *R* in the range cause tokens (elements) of type *D* in the domain on enough occasions (and in the right ways) a classifying disposition will come about such that a new token (element) of type *D* will trigger the disposition to classify the corresponding range token (element) as an *R*. On a projectivist account of color there are no red elements in the range that cause the appropriate domain elements. So even if there is a really good causal theory of how some classifying dispositions come about (I do not know of any), it cannot account for all of the ways in which classifying dispositions come about.

Consider how this account handles four cases.

(a) Imagine seeing a large box as yellow on a lawn that looks green such that one side of the box is in full sunlight and another side is in its own dark shadow, which extends to the lawn. The domain element of the 3-D visual model that represents the yellow box will be yellow. The domain element that represents the shadow will be black or gray. A shadowed surface of the box in the external world consists of two spatially overlapping range elements; a box and a shadow. These two elements correspond to two elements in the visual model; one is yellow (unmixed with black) and the other is black or gray (unmixed with yellow). Color constancy in spatially contiguous regions—seeing the box as uniformly yellow even though the sides of the box are illuminated differently, or seeing the shadow as black even though it falls on surfaces of different colors—is a consequence of the way elements in the domain of the mapping have an identity as elements in the domain. Again, a yellow element in the domain corresponds to an element in the range, which is seen as a yellow box; and a black element in the domain corresponds to an element in the range, which is seen as a black shadow. The visual model represents these two range elements as overlapping in space. There may also be a third white element in the domain that corresponds to an element in the range, which is seen as the lightness of the illumination of the scene and which also overlaps with the yellow box element. (Like shadows, seeing glares and objects behind colored transparencies involve seeing one element in the range behind another element in the range, and involve similar color constancy considerations.)

Color constancy is a consequence of domain element identity. Byrne and Hilbert assume that perception of color constancy is the result of computations that use information about the present illumination and its reflection from an observed surface to calculate the reflectance of the surface. The spectral reflectance of the sunny side and the shadowed side of the yellow box have the same spectral reflectance. I claim that if something like these computations takes place, which some vision scientists doubt, they are used by the visual system to construct the elements in the domain. Since the elements of the domain determine how the world is seen, the computations indirectly determine the apparent colors of range elements. (Like Byrne and Hilbert, Stephen E. Palmer (*op. cit.*) presupposes in his discussion of lightness constancy and color constancy of surfaces that the surfaces in question have already been identified, whereas I assume that these constancies are part of surface identification.)

(b) A person with color agnosia can pass all the tests for color blindness, which are color matching tests, but cannot identify the *color of an object that does not have a standard *color—ripe bananas have a standard *color but sweaters do not. In general, a color agnosic can classify things accurately for they correctly classify cars, pencils, and windows as cars, pencils, and windows. Furthermore, the problem is not due to a linguistic deficiency.[16] The color agnosic can match colors but cannot identify *colors. Why? An answer is suggested by the theory outlined above.

---

[16] A person, who can pass all the tests for color blindness but cannot identify the *color of an object that does not have a standard *color because she lacks the verbal links to *color terms, is not a color agnosic. Such a person is surprised when handed a blue banana, but a color agnosic is not surprised. (See Oliver Sacks, *An Anthropologist on Mars: Seven Paradoxical Tales* (New York: Knopf, 1995), p. 17.) For more on what is known about color agnosia, see R.L. De Valois and K.K. De Valois, "Neural Coding of Color," in Edward C. Carterette and Morton P. Friedman, eds., *Handbook of Perception*, Volume 5: *Seeing* (New York: Academic Press, 1975), pp. 117–66, reprinted in Byrne and Hilbert, eds., *Readings on Color: The Science of Color*, Volume 2 (Cambridge: MIT, 1997), pp. 93–140; M. Kinsbourne and E.K. Warrington, "Observations on Color

I begin with the distinction between color classification and color matching. In color matching, a perceiver is determining whether two color patches look the same or different in color. Color classification is a matter of seeing an object as having a specific color. For example, when one sees a tomato as red, one is not matching the color of the tomato to some color patch. Color classification implies color matching. If x and y each have the color C, then x and y have the same color (C, of course). The converse does not hold. By color matching alone, three distinct dimensions of color can be constructed: hue, saturation, and lightness.[17] However, the hues cannot be distinguished by color matching alone because that x is the same (or different) in hue from y does not imply that either x or y is red or blue or any other hue. So, although color matching is enough to generate the color solid from color patches, it does not have the resources to define what it is that the red parts of the color solid have in common.[18] (Clearly, one could define the *redness in the color solid by ostensive definition, but this goes beyond color matching.[19]) My point is that having color matching skills does not imply having color classification skills.

Since the color agnosic can match colors as well as a normal observer, we can assume that she has a visual system with the right dyadic disposition: if elements X and Y in the domain are the same color, say a shade of red, she will see x and y as being the-same-in-color. However, the color agnosic does not have a well working monadic disposition; although the domain element, X, is red and the subject sees x as similar to other objects in the range whose domain object is also red, she does not see x as a member of a class that extends beyond the range elements. The monadic disposition is duplicating the information in the dyadic disposition. This

---

Agnosia," *Journal of Neurology, Neurosurgery, and Psychiatry*, XXVII (1964): 296–99; and N. Geschwind and M. Fusillo, "Color-Naming Defects in Association with Alexia," *Archives of Neurology*, XV (1966): 137–46, reprinted in Byrne and Hilbert, eds., Volume 2, pp. 261–75. Geschwind and Fusillo show that children also have a color agnosia like problem. For more on the difficulties children have with color naming, see Jules Davidoff, *Cognition through Color* (Cambridge: MIT, 1991), pp. 148–53.

[17] The details of the construction can be found in Averill, "The Relational Nature of Color," pp. 557–60, and Austen Clark, *Sensory Qualities* (New York: Oxford, 1993).

[18] This is a well-known point in color science. For an example of an expert in color science making this point, see J. McCann, "Simultaneous Contrast and Color Constancy: Signatures of Human Image Processing," in Steven Davis, ed., *Color Perception: Philosophical, Psychological, Artistic, and Computational Perspectives* (New York: Oxford, 2000), pp. 88–89.

[19] Averill, in "The Relational Nature of Color," sets out in great detail a system for matching color patches and shows how the patches can be arranged to form the color solid. But where is yellow in this color solid? He claims (p. 562) that it is the job of the linguist to say what patches are yellow. On Averill's account, what yellow things have in common, in virtue of which they are yellow, is that we call them yellow. This is not illuminating.

would explain why a color agnosic appears to be guessing when asked the *colors of objects that do not have a standard *color. If this is right, the color agnosic sees objects by means of colors, just as normals do; however, she visually represents an object, whose domain element is red, as a member of a small subclass of the class a normal human would represent such an object as belonging to.

Before leaving color matching note that being able to draw color distinctions is the flip side of color matching. It is possible for there to be subjects who have an extraordinary ability to draw color distinctions: that is, they consistently see differences in color patches that most of us cannot distinguish. Contrast this ability with the ability to classify by *color. For the most part, *colors are defined for "normal observers" so there is no such thing as extraordinary perceivers of *colors. Indeed, there is no way to tell if one has an extraordinary ability to see things as colored that does not depend on color matching.

(c) It is possible to see a flash, low in intensity or in the periphery of the visual field, without seeing it as colored. More generally, it is possible to see motion without seeing it to have color. These are not cases where the domain element corresponding to the flash has no color. These are cases where color classification fails to take place. Unlike color agnosia, the subject has the appropriate classifying dispositions to see things as colored, but the colored domain element corresponding to the flash does not last long enough to trigger the color classifying disposition.

(d) There are many cases where one sees but is not conscious of seeing things. Walking is a visually guided task, but often when we are walking we are not paying attention to what we see. A person with blindsight can see some shapes and *colors in the "blind" part of her visual field—and so the person visually models and classifies. What such a person cannot do is attend to what she sees. In subliminal perception one is unable to pay attention. Hence, visual consciousness of what one sees requires attending, which is a level of visual processing that goes beyond the modeling and classifying involved in seeing. This point fits nicely with the claim made above that the sensuousness of c-colors exists independently of visual consciousness.[20]

II.4. Color Density. Since we only have R-G, B-Y, and W-B opponent cells, the firings of the opponent cells that make up a domain element can only have a few c-colors: c-red, c-green, c-blue, c-yellow, c-white, c-black, and c-gray (for base firing rates, which only occur when opponent cells are not stimulated).[21] Using the notion of color density, it was explained above how we are able to see orange and, by extension, the large number of colors in the color solid. Most of these colors are perceptual color mixtures. Note that saturation is a dimension in the color solid, where hues are more or less saturated with grays. Of course, this is just a special case of perceptual color mixture.[22] However, this still leaves some questions unanswered about the colors we see. How is it that there is only one c-white, yet there is both brightness and lightness? And how do the colors of transparent objects differ from the colors of opaque objects?

---

[20] See Palmer, pp. 531–70, for more on attention and visual consciousness.

[21] There may be colors other than those used in human vision. Presumably, tetrachromats use colors that humans do not (see E. Thompson, A. Palacios, and F. J. Varela, "Ways of Coloring," *Behavioral and Brain Sciences,* XV (1992): 1–26; and Thompson, "Novel Colors," *Philosophical Studies,* LXVIII (1992): 321–49). Short of becoming tetrachromats, there seems to be no way for us to have access to novel colors.

Consider the distinction between lightness and brightness. Lightness is attributed to objects that are seen as opaque nonradiant objects, and ranges from black to gray to an upper limit of pure white. Brightness is attributed to what is seen as radiation, which can vary from nonexistent to being too intense to look at. There is no black, gray, or brown radiation. (The distinction between lightness and brightness does not exist at the level of aperture colors.[23]) Although a television image is due to radiation, the colors in the picture can include black, gray, or brown because the picture is not seen as a radiating object. What is going on when our visual system represents a piece of paper, seen in bright sunlight, as white? On the account suggested here the element in the domain corresponding to the paper is white and the element in the domain corresponding to the glare is very white; that is, the density of white firings corresponding to the paper is less than the white firings corresponding to the glare. So the range element corresponding to the paper is seen as white (and it is classified as a color that has an upper limit) and the range element corresponding to the glare is seen as very bright radiation (and it is classified as a color that has no upper limit.) The distinction between lightness and brightness at the level of domain elements is a color density distinction; at the level of how we see the world it is a classification distinction.[24]

Color density has been used above to explain perceptual color mixtures, but there is another kind of case. If a domain element has a low density of c-green cells, and no other colors, we may assume that the corresponding range element is seen as transparent green. We can think of transparency as a case where a color is perceptually mixed with no color. Such perceptual mixtures would correspond to some paint mixtures. If one takes a red water color paint and mixes it with a lot of water one gets a washed out red. Such a washed out red corresponds to a domain element that has a relatively low density of R-G cells firing for red and no other cells in it have a color.

---

[22] This account of perceptual color mixture allows for the possibility that a range element might be seen as reddish-green if, in the corresponding domain element, the densities of R+G- cells firing above their base rate and those firing below their base rate were equal, and evenly distributed among one another. Some researchers claim to have seen reddish-greens; see V. Billock, G. Gleason, and B.Tsou, "Perception of Forbidden Colors in Retinally Stabilized Equiluminant Images: An Indication of Softwired Cortical Color Opponency?" *Journal of the Optical Society of America*, XVIII, 10 (2001): 2398–403.

[23] See Leo M. Hurvich, *Color Vision* (Sunderland, MA: Sinauer, 1981), p. 61.

[24] These considerations are easily generalized to include night vision, where objects look black, white, and various shades of gray.

II.5. Looks Red. A range element looks red to a subject when the corresponding domain element is red and is classified as red. The representational content of looking red, when looking at a ripe tomato, is that the tomato belongs to the class of things that have the property of being red. This is misrepresentation because tomatoes are not members of the class of red objects. R+G- cells are members of this class, when they are firing above their base firing rates. Does the experience of seeing a tomato as red also have a qualitative content?

I said above that R+G- opponent cells are correlated with red experiences when firing above their base firing rate, with green experiences when firing below their base firing rate, and conversely for G+R- cells. Okay, but how do we know which cells are R+G- cells? Well, they are assumed to be the cells that fire above the base firing rate when the eyes of the observer are stimulated by a *red light and fire below their base firing rate when stimulated by a *green light. However, we do not know that the cells in a subject that are fired by such lights are not G+R- cells. Indeed, it seems to be physically possible for *A* and *B* to be identical except that the R+G- and G+R- cells in *A*'s head correspond, respectively, to G+R- and R+G- cells in *B*'s head. If that were the case, then the R+G- and G+R- cells in *A*'s head would perform the same functions, respectively, that the G+R- and R+G- cells perform in *B*'s head. Both *A* and *B* describe a ripe tomato as *red, although it corresponds to a domain element in *A*'s head that is mostly c-red and to a domain element in *B*'s head that is mostly c-green. So a ripe tomato looks to *B* the way an unripe tomato looks to *A*. Hence, what it is like for *A* to see a ripe tomato is different from what it is like for *B* to see a ripe tomato. However, this qualitative content of *A*'s and *B*'s experience is not distinct from the representational content of their experience, because *A* visually represents a ripe tomato as red and *B* represents it as green.

When you read the above paragraphs keep this in mind: each individual should think of 'red' as referring to the property he or she visually attributes to *red objects seen under normal conditions, although there is no guarantee that each of us attributes the same property. To define '*red' for normal conditions, one points to a paradigmatic *red object under normal conditions and says "You are allowed to call that '*red', as well as anything that looks the same in color under normal conditions." When looking through a microscope or seeing fish under water, the definition of *red can be modified appropriately. Of course, in ordinary language we use 'red' for both red and *red.

I close with a brief comparison of my views with some of those set out by Ned Block in his "Inverted Earth" essay.[25] He and I agree that looking red has a representational (intentional) content and a qualitative, what-it-is-like, content. However, Block believes, and I do not, that these contents are distinct. Like Block, I think that the qualitative content cannot be captured by a public language because we cannot pick out the qualitative content of your experience of looking red and of my experience of looking red in a way that allows us to be sure that we are referring to the same property, as the possibility spectrum inversion makes clear. Block holds that the intentional (representational) content of looking red is given by "...its relation to red things, not its qualitative content" (*ibid.*, p. 55). Here Block assumes an objectivist position. I think that the representational content (and qualitative content) of looks red depends on its relation to red things, but these are R-G cell firings, not ripe tomatoes. Block holds that the representational content of looks red is expressed in public language. His claim depends on his assumption that there is an appropriate class of red objects in the objectivist sense, which allows for a strong distinction between is-red and looks-red. (Strong because it is possible for some things under normal conditions to look red but not be red, and some other things to be red but not look red.) I think that the objectivist cannot make out this distinction in a plausible way. I agree that there is a strong distinction between looks-red and is-red, but it is not a distinction that is useful for a public language. Public languages use *red, which only allows for a weak distinction between is-*red and looks-*red. Of course, *red is not to be confused with red. *Red specifies conditions on observers and viewing circumstances. When we see a tomato as red, we do not see it as red for such-and-such sorts of people under such-and-such lighting. We just see it as red. So, unlike Block, I hold that the color properties we visually represent objects as having are not the *color properties we linguistically represent objects as having.

EDWARD WILSON AVERILL

Texas Tech University

---

[25] In J. Tomberlin, ed., *Philosophical Perspectives*, Volume 4 (Atascadero, CA: Ridgeview, 1990), pp. 459–77.

LINEAR PERSPECTIVE AS A REALIST CONSTRAINT*

In Renaissance Florence, linear perspective was celebrated as the objective means to represent space convincingly. As such, it was soon accepted throughout Europe. Even if linear perspective is no longer a standard tool of the modern professional painter, the idea that linear perspective is required for a convincing pictorial space is still accepted in many circles. In times past, the Traditional Window Argument (based on the supposition that a painting of a scene is the projection of that scene onto the window pane through which it is supposedly seen) was taken to support the realist conviction that linear perspective is required for a convincing, faithful painting. Starting in the 1970s, Nelson Goodman and others began to question this privileged position of linear perspective along conventionalist lines: they held that there are various ways of pictorially representing reality, one of which uses linear perspective, and the choice between these ways is in the end based on convention. A painting using linear perspective is not—by itself—more faithful or convincing than a painting not using it. (I call a picture *faithful* when we experience the pictorial space in the picture as (closely) resembling the depicted world; I call a picture *convincing* when we experience the pictorial space depicted as (closely) resembling some intended world, real or imaginary.)[1]

---

* I thank Jan J. Koenderink for meticulously answering numerous questions, and M. Meijsing and an anonymous referee for their very helpful comments on an earlier version of this paper.

[1] Note that I do not claim a resemblance between the two-dimensional, flat picture (for example, made of a wooden frame, canvas, and paint) and the three-dimensional world depicted. The resemblance is between the depicted pictorial space, the virtual 3-D space which we see in many pictures (the experienced depth which has recently been studied by psychologists like Koenderink), and that part of the actual 3-D world which is depicted on the canvas of the painting. In this paper I take resemblance as a primitive notion, and our capacity to recognize resemblance as a primitive ability we develop over the years.

Let me explain why I need the term 'faithful'. It is to express a complicated relation. Due to size constancy, we may see two heads (in 3-D space) as the same size even if their occlusion sizes or retinal sizes are not the same at all. I call a painting "faithful" (in this respect) if we also see the corresponding heads in the pictorial space (that is, if we see these depicted heads) as having the same size. It is part of my argument that in some situations the painter will have to use the same size of canvas to create this experience in the viewer (namely, when he does not use perspective and there is little or no pictorial depth in the painting); in other situations when he uses linear perspective and/or cannot avoid pictorial depth, he will have to use different canvas sizes. The point of the term 'faithful' is to have a concept that indicates the similarity between the two cases—I experience the heads in the painting as being the same size (see them as having the same size) just as I experience them in the real 3-D world.

0022-362X/05/0205/235–58

Conventionalists grant that we experience a linear perspective painting as more convincing, but this is so—they add—because linear perspective has pervaded our Western pictorial tradition, and because this conception, rather than a nonperspectival one, is "inculcated," to use Goodman's happy phrase.[2] Linear perspective is not objectively required for fidelity, they argue; it is a convention of which we are not aware due to our Western upbringing.

Goodman is quite convincing in his critique of the Traditional Window Argument. Thus, the realist who wants to claim that linear perspective is one of the objective constraints on the creation of a faithful pictorial space needs another argument. E.H. Gombrich's ingenious defense of the Eyewitness Principle (based on the objective feature of the rectilinearity of light) is not enough, however; nor, I will argue, is the objectivist approach of John Hyman (based on the objective feature of projection) in terms of relative occlusion size.[3]

I will argue for the position that, to be faithful, a painting which successfully creates pictorial space by means of depth cues needs linear perspective to counteract the working of size constancy within that pictorial space. I conclude that linear perspective is one of the objective constraints on the creation of a faithful or convincing pictorial space.

I. THE TRADITIONAL WINDOW ARGUMENT

The *Traditional Window Argument* notes that a scene, as seen through a window with one eye, presents the same light pattern as a picture of the scene copied onto the window pane. The reflected light patterns are the same, and so our sensory input from the scene and from the picture of the scene must be (approximately) the same. (For the argument to go through, it should claim as well that our experiences are approximately the same.) That is to say, when a painter paints a scene as it is projected on the window pane, his painting will evoke approximately the same experience as the real scene and thus his painting (that is, its pictorial space) will look to a viewer like the scene painted: his painting will be a faithful picture of reality. It should be noted that the scene-as-projected-on-the-window will be projected there in accordance with the optical laws, and since linear perspective is based on these optical laws, the scene will in fact be painted in accordance with linear perspective. Conversely, if the painter deviates from linear perspective, he will cause a discrepancy between the way we experience the scene as seen through the window and the way we see the scene in the painting. Therefore, if a painter wants to depict some scene faithfully, he must employ linear perspective. Or in other words, for a faithful picture the use of linear perspective is required, and this necessity is nonconventional as it is based on projection and the optical laws of nature, expressed geometrically, that are the basis of linear perspective.

---

[2] *Languages of Art: An Approach to a Theory of Symbols* (Indianapolis: Hackett, 1976), p. 38.

[3] See Hyman, "Perspective," in David Cooper, ed., *A Companion to Aesthetics* (Cambridge: Blackwell, 1992), pp. 323–27, and "Pictorial Art and Visual Experience," *British Journal of Aesthetics*, XL (1980): 21–45.

Certainly, in order to create a pictorial space convincingly, a painter must also use other depth cues such as occlusion (the partial occlusion of some object by another object), light and shadow, relative textual density (defined as projected retinal density of a cluster of objects or textures as they recede into the distance), and aerial perspective (in the distance colors get hazier). But linear perspective must be one of the depth cues, and without it there is no faithful portrayal of depth in a picture. Linear perspective is thus not a convention but a necessary condition for the faithful depiction of pictorial space; or so it seems.

In this analysis, linear perspective has two closely related aspects: (1) the convergence aspect (parallel-lines-in-reality appear on the picture surface as lines converging to a vanishing point), and (2) the reduced-size aspect (the relative reduction of the depicted sizes of the objects is determined by the relevant converging lines just mentioned, even if these lines are not drawn on the picture). Both aspects will play an important part in my argument.

II. THE CONVENTIONALIST CRITIQUE

Goodman is not impressed by these arguments. In his *Languages of Art: An Approach to a Theory of Symbols* (*op. cit.*), he presents a number of arguments why linear perspective is not required for a faithful, spatial picture. I take Goodman as my starting point, as recent critics have not improved on his critiques. I will discuss here two of his main arguments.[4]

His main *theoretical critique* revolves around the insight that, due to the different observational conditions under which the painting and the scene are observed, the painting is a translation which has to be read, and thus a painting that looks faithful to us is one which uses our familiar system of representation. Goodman argues that the painting and the scene are observed under very different conditions. The painter would be well advised to take these conditions into account. He has to discover "what light rays, under gallery conditions, will succeed in rendering what he sees. This is not a matter of copying but of conveying." It is a "translation of a sort." One consequence of this is that "pictures in perspective, like any other, have to be read" (*op. cit.*, p. 14). In our modern Western world, we do not notice the need for an interpretation: the reading of pictures in perspective has become a "virtually automatic habit" (*op. cit.*, p. 38). Perspective is a convention, but one that is hard to recognize as such. We experience a painting as faithful, not because the painter uses some particular "faithfulness-creating" system of representation (for example, linear perspective), but because "the system of representation employed in the picture [is our] standard system" (*op. cit.*, p. 38). That is, whether a painting looks to us a faithful or convincing picture—or in Goodman's terms, a representation with *fidelity*—does not depend on some objective relation between the picture and the depicted scene, but on the inculcation of a particular convention in the perceiver. The facility of the interpretation makes us overlook the inevitable presence of the interpretation.

---

[4] I will discuss Goodman's criticism that the Traditional Window Argument is mistakenly based on the assumption that we look at the painting with one eye later in the context of the "Remaining Worry" (section XI).

Marx Wartosky[5] takes the same line, and so do postmodernists like Murray Krieger[6] and Norman Turner,[7] and the psychologist Margaret Hagen.[8] Once we break the spell, we can experience other conventions as equally natural.[9] We make nature as much as we make the painting. The kind of perspective that has become standard in painting has also become part of reality, that is, part of our everyday world version.

Goodman also offers a *practical*—and it seems equally fatal—*critique* of the Traditional Window Argument: even to the Western inculcated viewer some pictures in linear perspective do not look faithful.

> The photograph of a man with his feet thrust forward looks distorted, and Pike's Peak dwindles dismally in a snapshot. As the saying goes, there is nothing like a camera to make a molehill out of a mountain (*op. cit.*, p. 15).

I include a photograph of my young friend Jonathan with his hand thrust out to illustrate the point made.

In spite of the use of linear perspective, the photograph (figure 1) does not look faithful: the hand looks too big. Goodman concludes: "[linear] perspective provides no absolute or independent standard of fidelity" (*op. cit.*, p. 19). It is no more than a well-used convention within

---

[5] "Rules and Representation: The Virtues of Constancy and Fidelity Put in Perspective," *Erkenntnis*, XII (1978): 17–36.

[6] "The Ambiguities of Representation and Illusion: An E.H. Gombrich Retrospective," *Critical Inquiry*, XI (1984): 181–94; cf. "Western naturalism [is] just another—and different—code" (p. 191).

[7] "Some Questions about E.H. Gombrich on Perspective," *The Journal of Aesthetics and Art Criticism*, L (1992): 139–50; cf. "perspective has become the dominant pictorial convention of our place and time" (p. 149), and "The abstractions of projective geometry are attributed to things" (p. 145).

[8] *Varieties of Realism: Geometries of Representational Art* (New York: Cambridge, 1986).

[9] Rudolf Arnheim made a similar point: we need only a "shift of level...to make the Picassos, the Braques, or the Klees look exactly like the things they represent"—"What Looks Lifelike?" in his *Art and Visual Perception: A Psychology of the Creative Eye* (Berkeley: California UP, 1954/1974 revised edition), p. 19. And Hagen agrees: "Arnheim vowed that with familiarity, the multiple station point images of the cubists could come to seem realistic to us. They will" (p. 225).

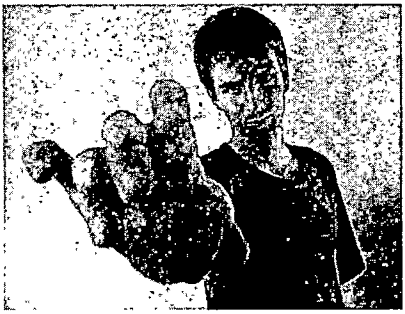


Figure 1

the Western tradition of representing reality, which does not guarantee a faithful representation even to those who have grown up within that tradition. Faithfulness of a picture is "relative, determined by the system of representation standard for a given culture or person at a given time" (*op. cit.*, p. 37).

The realist who wants to defend the nonconventional significance of linear perspective for a faithful, convincing picture, had—so it seems—better look for another argument than the Traditional Window Argument. Both the Theoretical and the Practical Argument seem quite devastating.

III. GOMBRICH'S EYEWITNESS PRINCIPLE

Gombrich resisted the idea that linear perspective is merely a convention. Yet, in contrast to the Traditional Window Argument, he stressed that our subjective visual experience is not determined by the physical or optical world. "There is no fixed correlation between the optical world and the world of our visual experience."[10] However, Gombrich is firm on his insistence that this does not turn linear perspective into a convention.

Gombrich defends his position with the aid of the Eyewitness Principle. The principle requires that "the artist must not include in his image anything an eye-witness could not have seen from a particular point at a particular time."[11] And you need linear perspective "if you want to map precisely what anyone could see from a given point" (*ibid.*, p. 281). As all this turns on the objective fact of occlusion, which in its turn is based on the rectilinearity of light rays, linear perspective is therefore independent of convention.

However, this defense of linear perspective against conventionalists is faced with two serious difficulties: first, the Eyewitness Principle does not single out linear perspective as the only possibility. For instance, there are all kinds of deforming perspectives which show what can be seen, even if they do not preserve the shape.[12] Yet they successfully map everything that one could see from a given point. Second, and equally worrisome, Gombrich's defense is not concerned with how things will *appear* to us at a certain point; it merely tells us what features someone could see from a given point.[13] Thus it leaves unexplained why linear perspective at least seems to play such a crucial role in the creation of the convincing appearance of a pictorial space.

---

[10] "Mirror and Map: Theories of Pictorial Representation" (1978), in his *The Image and the Eye: Further Studies in the Psychology of Pictorial Representation* (London: Phaidon, 1994), pp. 172–214, on p. 178.

[11] "Standards of Truths: The Arrested Image and the Moving Eye" (1976), in Gombrich, pp. 244–77, on p. 253.

I conclude that both the Eyewitness Principle and the related idea of occlusion achieve too little. I will now look at a more recent position that heavily leans on occlusion, namely, Hyman's analysis in terms of relative occlusion sizes. I shall argue that it also fails for related reasons.

IV. RECENT OCCLUSION THEORIES

In his 1992 article "Perspective" (*op. cit.*), Hyman applies the occlusion analysis to the issue of conventionality of linear perspective, or as he calls it, of *artificial perspective.* I will therefore concentrate my discussion on his arguments, disregarding the related position of Robert Hopkins.[14]

Hyman notes that occlusion shape ("an object's outline or silhouette," or what it excludes from the observer's view) and relative occlusion sizes are "visible properties of the physical world" and not "features of our subjective visual impressions" (*op. cit.*, p. 325), as, for example, Christopher Peacocke maintains.[15] This is clear from the fact that "I can mistake the relative occlusion size of two trees, and my mistake can be corrected by measurement and geometrical calculation" (*op. cit.*, p. 325). Once we realize this, we can, so Hyman claims, see that linear perspective "is an objective technique for depicting non-planar spatial relations" (*op. cit.*, p. 326), like overlapping and foreshortening.

> Perspective diminution [that is, diminution in line with linear perspective] allows the painter to depict relative distance from the spectator, by making the relative size of the parts of the picture which depicts various objects the same as the objects' relative occlusion size (*op. cit.*, p. 326).

---

[12] For example, curvilinear perspective would depict everything that one could see from a given point, even if it might not appear that way to the human observer. Carel Fabritius gives an existence proof of this in his *View on Delft* (in London at the National Gallery).

[13] See Gombrich, "Image and Code: Scope and Limits of Conventionality in Pictorial Representation" (1978), in Gombrich, pp. 278–98, on p. 281.

[14] Hopkins speaks about outline shape in his *Picture, Image, and Experience* (New York: Cambridge, 1998). As Hyman discusses the issue of perspective specifically, I will concentrate my discussion on his arguments.

[15] "Depiction," *Philosophical Review*, xcvi (1987): 383–410.

Perspective diminution, and thus linear perspective, is objective as it results from the application of projection and is based on relative occlusion sizes which are objective. It is not conventional.

Hyman is clearly right in his claim that relative occlusion sizes are objective. We should ask two questions, however. The first is an empirical question: Do relative occlusion sizes in fact generate a convincing picture? We have at least two reasons to doubt this. First, there is Goodman's Practical Argument: there are many pictures in linear perspective (and hence pictures using relative occlusion sizes) that do not look convincing. And second, so long as we have not succeeded in creating some pictorial depth, the strategy of relative occlusion sizes will not yield a convincing picture. I will explain and defend this claim in the next section, when I discuss size constancy.

Even if we grant that, *once* the picture has pictorial depth, relative occlusion sizes fit right in and help to create a convincing picture, we want to know—and this is the second question—why relative occlusion sizes succeed in doing so. Is the painting faithful because linear perspective (with its relative occlusion sizes) is our well-inculcated convention (as Goodman argues)? Or is the painting faithful because the relative occlusion sizes used in the painting correspond to the objective occlusion sizes of the things in the world (as Hyman claims)? Or is the painting faithful because linear perspective (and so relative occlusion sizes as well) is required due to some feature of our perceptual apparatus (as I will claim)? So I conclude that relative occlusion sizes surely are objective features, but that the mere mention of this objectivity does not answer the conventionalist objections.[16] We have to look closer at size constancy, and what it does to our perception of real space and of pictorial space, to get a more satisfactory answer.

V. THE CONSTANCY ARGUMENT

In a recent handbook on perceptual constancy, S.P. McKee and H.S. Smallman describe size constancy as follows: "the tendency for objects to appear much the same size over a wide range of distances in spite of the changes of the retinal images [occlusion sizes] associated with distance of the object."[17] We can "observe" size constancy scaling[18] (the operation which realizes size constancy) when we hold one hand twice as far from

---

[16] Cf. my "Occlusion Shapes and Sizes in a Theory of Depiction," *British Journal of Aesthetics*, XLIV (2004): 319–41.

[17] "Size and Speed Constancy," in Vincent Walsh and Janusz Kulikowski, eds., *Perceptual Constancy: Why Things Look as They Do* (New York: Cambridge, 1998), pp. 373–400, on p. 373.

[18] Richard L. Gregory rightly distinguishes between constancy as a result and constancy scaling as an operation which realizes constancy as a result—see his *Eye and Brain: The Psychology of Seeing* (Princeton: University Press, 1997, 5th edition).

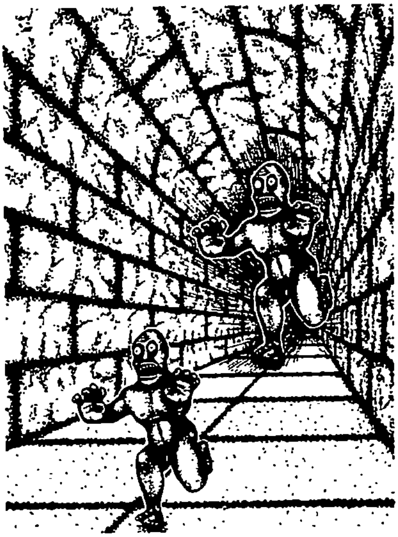


Figure 2

our eyes than the other one. In spite of the fact that their occlusion sizes differ quite a bit, the hands still look about the same size.

In the 1930s, Robert Thouless found that, when there are clear depth cues, size constancy is fairly perfect within a large range. These results are now widely accepted, although some haziness persists about the details. But if this is how the world appears to us (that is, we see equal-sized objects as about equally large), why does a painter, aiming at a faithful or convincing appearance, not represent reality as it appears to him, constancy and all? Why does he not paint equal-sized objects which look equally large, as equally large? This suggestion makes a cardinal mistake: it neglects the fact that within the virtual space of the painting there is a further size constancy to be compensated. A painting is not just a canvas, but also a presentation of a virtual scene in a virtual space, which itself is susceptible to the operations of our perceptual apparatus. Both Gombrich and Richard L. Gregory (*ibid.*) mention this phenomenon. It can be shown by means of the following two figures. A simple picture created by Roger N. Shepard (figure 2) demonstrates that constancy scaling works in pictures with depth.

What happens? Apparently, our brain reacts to the pictorial depth of the picture, notes the equal occlusive sizes of the depicted monsters, and subsequently "calculates" the true sizes of the monsters. Our brain "concludes" that the monster in the distance must be larger than the monster close by, and so we see it larger. This is constancy scaling at work within the picture. If we remove the depth cues (see figure 3), the impression changes. The difference in experienced size strongly diminishes—unless we

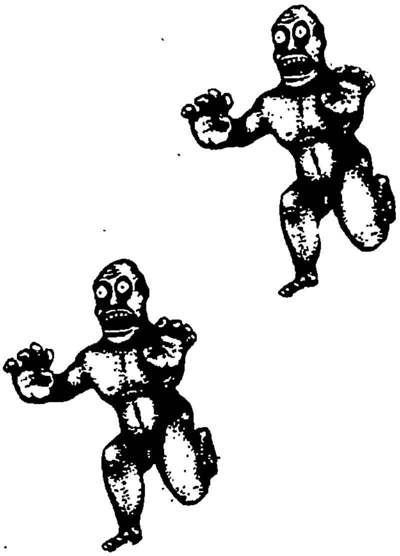


Figure 3

create our own pictorial depth, for example, by remembering the previous picture or by imagining the monsters as running in the snow.

If this analysis is correct, effective depth cues in a painting have a frustrating effect. Suppose we want to paint two equal-sized objects, one close by and the other at some distance from us. In spite of their different distances from us, and in spite of their different occlusion sizes, these objects look to us to be approximately the same size. (This is size constancy scaling in the real world.) Suppose that, due to this impression, we were to paint them with approximately the same physical size on the canvas, so that the pieces of canvas they cover have about the same size. That is, we paint something like figure 3. Suppose further that we succeed in creating a convincing pictorial space with effective depth cues. (For example, we add the corridor of figure 2). We are now in for a nasty surprise. In the painting, the object at a distance will look larger than the object closer by. It looks larger due to the operation of constancy scaling within the virtual space of the painting. Roughly, the more impressive the pictorial space is in a picture, the more the picture will be affected by constancy scaling, and the more the increase in apparent size has to be compensated for.

It is the reverse problem, the absence of constancy scaling in a picture without pictorial depth, that is the undoing of Hyman's analysis in terms of relative occlusion size: if there is no (convincing) pictorial depth in the picture, then there is no constancy scaling in that picture requiring compensation, and relative occlusion sizes will look wrong. Let me elaborate briefly. Suppose that in figure 2 we had drawn the two monsters according to their relative occlusion sizes in reality—that is, we had drawn the distant monster as smaller as the near monster. If we now remove the depth cues so that within the picture size constancy scaling does not operate, the smaller canvas size of the further monster will not be compensated for by constancy scaling. So the more distant monster-in-the-picture will look smaller than the near monster-in-the-picture, even though the monsters would look the same size in real, visual space (due to constancy scaling in that space). Consequently, the monsters in the picture will not look like the real monsters as they would appear to us in reality, and so the picture is not faithful. Thus, relative occlusion size is not enough for a convincing depiction of objects. We first need a convincing pictorial depth.

Pictorial space, then, is not without its dangers. The constancy scaling which operates in pictorial space, has to be compensated for. Within the pictorial space of figure 2, the size of the distant monster has to be diminished to obtain a convincing impression of two equal-sized monsters. How much should we diminish this monster in the distance? This is where linear perspective comes in. We can see straightaway that reverse and parallel perspective will not accomplish the job; they do not diminish the size of the monster further away in pictorial space. But does linear perspective decrease the size appropriately?

Suppose I depict two equal-sized objects *A* (close by) and *B* (at some distance) as I (naively) experience them—that is, my depiction includes constancy scaling as it operates in real space. Since I experience *A* and *B* as equally large, I depict them equally large (the pieces of canvas covered by paint are of equal size), and I do so even if *B*'s occlusion size is twice as small as that of *A*.[19] Suppose further that someone helps me to create a pictorial space. Then the objects-in-the-pictorial-space (that is, *A* and *B* as they appear in the painting) will be affected by constancy scaling as it operates in pictorial space. This constancy scaling will enlarge *B*-far-in-pictorial-space as compared to *A*-nearby-in-pictorial-space. Therefore, since the pictures of *A* and *B* are physically the same size (they have the same paint-size), *B*-in-pictorial-space will look larger than *A*-in-pictorial-space. Remember, however, that in reality *A* and *B* look the same size (as a consequence of constancy scaling in real space). I will therefore have to reduce the picture (the canvas size) of *B*. How much should I reduce the picture of *B*? My task is to cancel the enlargement due to the constancy scaling in pictorial space.

Suppose now that my assistant creates such a convincing pictorial space that the constancy-scaling-in-pictorial-space and constancy-scaling-in-real-space have the same strength. Due to the constancy scaling within the pictorial space, *B*-in-the-pictorial-space will now look twice as large as *A*-in-the-pictorial-space.[20] I will therefore have to reduce the picture of *B* to half its (physical) size. With this reduction—and this is important—the sizes of the pictures of *A* and of *B* obtain the same proportion as *A*'s and *B*'s occlusion sizes (2:1), which in its turn corresponds to the sizes of *A* and *B* according to linear perspective. That is, if the virtual space within the painting has been painted so well that the constancy-scaling-in-real-space and the constancy-scaling-in-pictorial-space have the same strength, then by depicting objects according to their relative occlusion sizes (that is, by making use of linear perspective) I will precisely cancel the operation of constancy-scaling-in-pictorial-space. The *A*-in-the-picture and the *B*-in-the-picture will look the same size as the *A* and *B* in real space. So the resulting picture will be experienced as faithful (at least in this respect). This is the *Constancy Argument.*[21]

---

[19] The factor of two tells us how strongly the constancy-in-real-space affects the way in which the painter sees *B*.

Note that this argument explains why linear perspective (here in its aspect of reducing the depicted sizes of objects further away) is required for a convincing spatial picture: it is required to compensate for the constancy scaling within the pictorial space. If correct, the argument would also specify the objectivity of linear perspective against the conventionalist position; the necessary use of linear perspective to obtain a faithful (spatial) picture does not depend on the inculcation of a linear perspective tradition but on two features of our perceptual apparatus: namely, size constancy scaling within visual space and size constancy scaling within pictorial space.

However, note as well that the argument clearly depends on an empirical assumption, the *Equal Constancy Assumption*: the spectator's size constancy scaling works to the same extent in real space as in the virtual space of the picture. This assumption is all but self-evident.

Whatever the tenability of that assumption, the argument already has an interesting conclusion: if there is some constancy scaling in pictorial space, and the first picture with the two monsters (figure 2) demonstrates that there is, then some reduction is required. This *Weak Constancy Argument* refutes Goodman's claim that linear perspective, reversed perspective, and parallel perspective are just three different conventions. Neither reversed perspective nor parallel perspective diminish the objects in pictorial space, so they cannot possibly compensate for the constancy at work in that pictorial space. (I realize that the counterexamples of the conventionalist's Practical Critique should still impede any triumphalism.)

---

[20] Remember that we had assumed that, for object *B*, two was the strength of the constancy-scaling-in-real-space. We just supposed that the constancy-scaling-in-pictorial-space was equally strong.

[21] This argument does not depend on the easy numbers used and can be generalized.

VI. EQUAL CONSTANCY ASSUMPTION

Do people have the same degree of constancy scaling in real, visual space and in pictorial space? Well, how effective is size constancy scaling in real space? Gregory summarizes Thouless's results: "Constancy was generally almost perfect for fairly near objects, though it broke down for distant objects, which can look like toys" (*op. cit.*, p. 223). The recent handbook article by McKee and Smallman (*op. cit.*) claims accurate size constancy over a range from 10 to 120 feet, when adequate information about depth [is] available. However, over the years it has also become clear that "perfect constancy [is] a rare occurrence."[22] What everybody agrees on is that size constancy improves if there are more and better depth cues.[23] Size constancy in real space is under appropriate conditions quite impressive,[24] even if over almost the whole range there is some regression to the relative occlusion sizes.[25]

About size constancy in pictorial space, there is, to my knowledge, preciously little information.[26] The adjusted two-monsters picture (figure 4) shows that there is at least some degree of size constancy within pictorial space, and that the converging lines of linear perspective on their own (even in this imprecise version) are sufficient to produce some size constancy scaling.

But there are good *prima facie* reasons for doubting that constancy scaling operates as effectively in pictorial space as in real space. First, following Gregory's intuition, we should realize that "the artist can hardly hope to provide all the depth cues available from reality" (*op. cit.*, p. 185) and that therefore the observer's constancy scaling in pictorial space, affected as it is by the limitation of depth cues, will—probably—work less well than in real space.

Second, we may surmise that in pictorial space there may well be a greater regression to the relative occlusion sizes (here, the surface sizes) due to the greater ease with which the surface sizes will impose themselves

---

[22] William Epstein, "Historical Introduction to the Constancies," in Epstein, ed., *Stability and Constancy in Visual Perception: Mechanisms and Processes* (New York: Wiley, 1975), pp. 1-22, here p. 19.

[23] Cf. McKee and Smallman, p. 393.

[24] For a very different view, see the Gibsonian I. Gordon, "Gombrich and the Psychology of Visual Perception," in Richard Woodfield, ed., *Gombrich on Art and Psychology* (Manchester: University Press, 1996), pp. 60–77.

[25] Suppose that *A* and *B* are equal-sized objects, and that from observer *O*'s viewpoint *A*'s occlusion size is larger that *B*'s occlusion size. When there is complete size constancy scaling, *A* and *B* will have equal apparent visual sizes for observer *O*. There is *regression to the relative occlusion sizes* (that is, there is less than complete size constancy scaling) when for *O* the apparent visual size of *A* is (somewhat) larger than that of *B*.

[26] In her article, "Perceiving Pictorial Space," Sheena Rogers does not even mention the topic of size constancy—in Epstein and Rogers, eds., *Perception of Space and Motion* (San Diego: Academic, 1995), pp. 119–63.

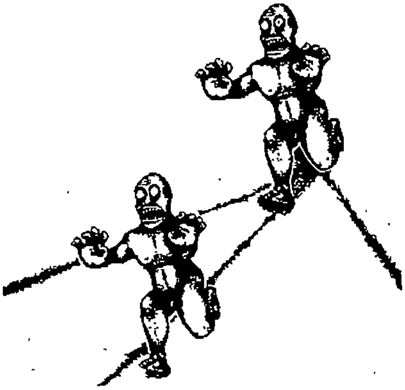


Figure 4

upon us in pictorial space. It is the close proximity on the flat painted surface itself which (easily) leads to the interference of the occlusion sizes with size constancy. This is likely the explanation as to why, in Vermeer's wonderful painting *The Soldier and the Laughing Girl*, the girl is often experienced by viewers as too small. On the painting's surface, the soldier and the girl are placed so close together that, for many, this leads to an uncomfortably strong influence of the occlusion (surface) size. The size constancy scaling which would have operated in real life had we faced the soldier and the girl, does not—so we find—operate strongly enough in Vermeer's pictorial space. The occlusion sizes (the surface sizes) can thus have a frustrating influence of the operation of size constancy within pictorial space. I will return to this issue of regression to the occlusion sizes later.

I tentatively conclude that (1) there is (a fair amount of) constancy scaling within pictorial space, that (2) the constancy scaling is dependent on depth cues within the picture, and that (3) the regression to the occlusion size will probably be larger than in real space.

These considerations lead me to suggest that the Equal Constancy Assumption is probably mistaken, though it is unclear to what extent. That is, we cannot use the Constancy Argument as it stands to argue for the nonconventionality of linear perspective itself. However, we have also seen that there is at least some size constancy in pictures with pictorial space. This constancy scaling is by itself enough to remove reverse perspective and parallel perspective as competitors. I called this Weak Constancy Argument "weak" because it does not exclude other size-reducing perspectives (other than linear perspective). Yet it does undermine the conventionalist claim that any kind of perspective might have given us faithful and convincing pictures, had we been brought up with it.

VII. A REDUCED PERSPECTIVE AS A SOLUTION?

Within pictorial space, constancy scaling operates. As a consequence, the painter has to reduce the objects more distant in pictorial space. The Constancy Argument offered a reduction following linear perspective as the proper one. However, if we assume that the size constancy scaling within a painting is less than in real space, the painter should perhaps "use a reduced perspective for maximum realism," or so Gregory (*op. cit.*, p. 185) suggests.[27] For instance, in our two-monsters picture with converging lines (figure 4), we should reduce the distant monster less than linear perspective would prescribe. Is this a viable alternative?

Both H.A. Sedgwick's *horizon-ratio relation*[28] and Gombrich's Eyewitness Principle testify against it. The horizon-ratio relation is based on the insight—well known in the Italian Renaissance—that the horizon is at the height of the painter's eye and that everything of such a height, viewed through the painter's eyes, will just reach the horizon.[29] Objects with half their length above the horizon are therefore twice as large as the height of the painter's eyes. The ratio between the total length of an object depicted and the length of that object underneath the horizon thus gives information about the relative size of the object compared to the size of the painter, and also—and very importantly—compared to the relative size of other objects in the picture. A painting using a "reduced perspective" would thus give the wrong information about the objects depicted. However, this argument concerns a mathematical relation which may be used as information,[30] but it does not necessarily tell us what we see or experience. Therefore, its relevance here is unclear.

The Eyewitness Principle brings us—in the present context—closer to the look of things. Suppose that in the two-monsters picture, the one with the converging lines (figure 4), we want to present the monsters as having equal physical size, and suppose we diminish the picture of the distant monster according to reduced perspective. That is, we diminish it, but not quite as much as linear perspective would specify. The complication is now that if we were to restrict our operation to the distant monster alone, he would cover some things in the background we can see in real space. To show *everything* that can be seen (which is the point of the

---

[27] The term 'reduced' is somewhat misleading, as the reduced perspective reduces less than linear perspective. It reduces the reduction of linear perspective, while parallel perspective does not reduce at all.

[28] See "The Geometry of Spatial Layout in Pictorial Representation," in Hagen, ed., *The Perception of Pictures*, Volume I (New York: Academic, 1980), pp. 33–89, and "Space Perception," in K.R. Boff, L. Kaufmann, and J.P. Thomas, eds., *Handbook of Perception and Human Performance*, Volume I (New York: Wiley, 1986), pp. 21–57.

[29] The plane between painter and horizon should be flat.

[30] Cf. Rogers, pp. 125ff., who provides some experimental evidence that people react to this informational invariant.

Eyewitness Principle), the painter, working on the restricted space of the canvas, would have to compress things around the distant monster. But we may now expect these compressed things to look too small. After all, in comparison to linear perspective, we increased the distant monster (in the picture) to make it look the right size, and now we have reduced (rather than increased) things around and behind the distant monster. The painting will not look faithful. Thus, given the demands of the Eyewitness Principle plus the effects of constancy scaling within pictorial space, reduced perspective leads to an unfaithful picture.

It might well seem that now, apart from linear perspective, all kinds of perspectives have been excluded,[31] and that thus the unique status of linear perspective as the necessary device for a faithful or convincing picture has been established. However, we should wonder how it is *possible* that linear perspective could lead to faithful pictures as its effectiveness is premised on the *prima facie* incorrect Equal Constancy Assumption. We have seen two reasons to question this assumption: first, there are fewer depth cues in a painting than in the real world, and, second, the occlusion sizes on the painting may well impose themselves on the brain to such an extent that they undercut the effectiveness of the constancy scaling within the pictorial space (that is, they cause a strong regression to the occlusion sizes). I will argue that (1) linear perspective in its convergence aspect answers the first doubt, and that (2) Goodman's counterexamples of his Practical Critique in fact demonstrate the limited power of the occlusion sizes in most linear perspective pictures.

VIII. LINEAR PERSPECTIVE IN ITS CONVERGENCE ASPECT AS A STRONG PICTORIAL DEPTH CUE

So far, I have concentrated on the reduced-size element of linear perspective. Constancy scaling required size reduction and linear perspective seemed to offer a promising reduction, yielding the Weak Constancy Argument. But the fact that there are fewer depth cues in a painting than in the real world casts serious doubt on the Equal Constancy Assumption of the Constancy Argument. Therefore, I will approach linear perspective from a different direction. I now look at its convergence element: the element of converging lines with a vanishing point. Figure 4 should impress on us that converging lines on their own already give a strong impression of pictorial space. So let us see what power the converging lines and the vanishing point have on their own. I shall argue that linear perspective in its convergent aspect alone is an effective creator of pictorial space, even when many depth cues from the real world are lacking.[32]

---

[31] Actually, it may be thought that so far curvilinear perspective has not been excluded from consideration as it is not your prototypical kind of reduced perspective. However—whether we want to call curvilinear perspective a reduced perspective or not—the last part of the argument above applies: in curvilinear perspective, some items are compressed relative to other items and we may again expect, for the reasons given, that these compressed items will look too small. Moreover, and this is a practical point, there would be not any reason to introduce curvilinear perspective in a painting when the painter aims at faithful appearances. Even if our appearances would have some curvature, it would be so small that on a painting it would not show: "perceptual spaces are sufficiently close to being Euclidian that little of practical import is gained by considering Riemannian curvatures"—James E. Cutting and Peter M. Vishton, "Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth," in Epstein and Rogers, eds., pp. 71–118, on p. 73.

For this purpose, the classic 1958 experiment by O.W. Smith, P.C. Smith, and D. Hubbard on perceived distance as a function of the method of representing perspective is highly relevant.[33] It provides experimental support for the critical and independent role of converging lines with a vanishing point. The experiment compared the different judgments of depth made by observers in the case of pictures with ever increasing depth cues: (a) a line drawing of a corridor (viewed head-on) without detail or shading, with only the converging lines and the mere indication of the end wall of the corridor, (b) a shaded (perspective) drawing of the corridor, (c) a line drawing of the corridor with some details beyond the converging lines, (d) a line drawing of the same corridor with full details, and (e) a photograph of the corridor. The authors assumed that the "errors of judgments of distances portrayed by the drawings should decrease as [the drawings] are made to resemble more and more closely real scenes by inclusion of shading and other details" (*ibid.*, p. 664). They found, however, that "there is no evidence that depth perceived in perspective line drawings [including the one with only the converging lines] differs from that of photographs" [with all the details] (*ibid.*, p. 674); and "since neither detail nor shading changed [the judgment of depth], the single most important factor in determining [these] judgments, under the conditions of this experiment, is that of the [linear] perspective" (*ibid.*, p. 673).[34] The value of this conclusion is still fully appreciated. Sedgwick refers to this research when he concludes that "the linear perspective information in the images was primarily responsible for the perception of distance."[35]

---

[32] There is actually some empirical support for making the distinction between these two aspects of linear perspective. R.N. Haber and Ch.A. Levin defend the claim that "theorists have failed to recognize that distance perception and size perception ...serve very different perceptual demands [and that]...distance perception and size perception...depend on largely non-overlapping sources of information"—"The Independence of Size Perception and Distance Perception," *Perception and Psychophysics*, LXIII, 7 (2001): 1140–52, on p. 1140.

[33] "Perceived Distance as a Function of the Method of Representing Perspective," *American Journal of Psychology*, LXXI (1958): 662–74.

[34] This result is even more impressive when we combine it with the result of another classic experiment by Smith and Smith, "Ball Throwing Responses to Photographically Portrayed Targets," *Journal of Experimental Psychology*, LXII (1961): 223–33. It turned out that subjects were equally good at throwing a ball at a target whether they saw the target in a real corridor, or whether they—monocularly—saw the target and the corridor on a (linear perspective) photograph of the corridor.

All this is not, of course, to say that linear perspective as convergence is the only source of information for depth in pictures. It is, however, a pivotal source. More or less on its own, linear perspective convergence, being "primarily responsible for the perception of distance," produces a forceful pictorial space, that is, a space in which constancy scaling apparently works effectively, in spite of the absence of many depth cues. This would explain why many linear perspective pictures impress us as faithful, even if they lack many real-life depth cues. The empirical evidence seems to indicate that a painter, in using linear perspective convergence, uses a means to create a forceful pictorial depth in which size constancy scaling plus the usual ambiguity we are used to are sufficient to create a convincing spatial picture, in spite of the fact that many depth cues from the real world are lacking. Call this the *Depth Argument.* Together with the Constancy Argument, this uniquely picks out linear perspective as an objective means to a convincing and faithful picture, or so it seems. It is an objective, nonconventional means as its successful operation is predicated on objective features of the human perceptual apparatus that codetermine the visual field of the perceiver: (1) its measure of constancy scaling in real space, (2) its measure of constancy scaling within pictorial space, and (3) its reaction to linear perspective convergence as a strong pictorial space cue.

Before we may draw this conclusion, however, we have to face two important criticisms and a remaining worry. The first criticism is the conventionalist's critique that my arguments depend on psychological data and that these data have been collected with the help of only Western subjects who have been inculcated by linear perspective. Call this the *Ultimate Conventionalist Argument.* The second criticism is Goodman's *Practical Critique* that, even to the Western inculcated person, some pictures in linear perspective do not look convincing. This may seem to cast serious doubts on the alleged power of linear perspective in its convergence aspect. The *Remaining Worry* concerns the quality of the pictorial depth which linear perspective in its convergence aspect produces: How could the resulting pictorial depth generate strong enough a constancy scaling that reduction according to linear perspective (that is, using relative occlusion sizes) yields a faithful or convincing picture? After all—a point made by Goodman—we look at a picture with two eyes, while the idea of linear perspective is based on looking from the station point with one eye. In real life, we see more depth with two eyes, while some pictorial depth is lost in our experience when we look at the picture binocularly.[36] A greater regression to the occlusion (surface) sizes, that is, a less perfect size constancy, is to be expected in the case of pictorial space.

---

[35] "Space Perception," p. 39.

I will argue that (1) the Ultimate Conventionalist Argument is just mistaken, that (2) Goodman's Practical Critique may be turned around to our advantage, and that (3) there is quite some flexibility in our perception and diversity in the world which fills the slack.

IX. THE ULTIMATE CONVENTIONALIST ARGUMENT

My arguments thus far have used a good amount of data about our perceptual system. The conventionalist must be infuriated by so much partiality. The data have been collected in the Western world! Small wonder then that linear perspective seems objective. Its conventionality has long since become "natural." When we look at the ethnographic literature, so the conventionalist tells us, we discover very different data. In W. Hudson's famous experiments, many Bantu people failed to see depth in his linear perspective drawings.[37] And, reusing Hudson's drawings, his results have been replicated many times. That is, ethnographic research seems to show that the reading of representations in linear perspective has to be learned. A representation in linear perspective may seem natural (nonconventional) to us; it is not. It all depends on the system of representation that has been inculcated. Thus, the conventionalists present an empirical argument to back up their position. This is the *Ultimate Conventionalist Argument*. It may well seem deadly but there are empirical considerations which strongly tell against it.

First, the history of painting is in stark conflict with this conventionalist analysis. When, in Renaissance Italy, linear perspective was first developed, the new kind of painting was experienced as more faithful even though no one had grown up in a tradition of linear perspective. And the same is true for Northern Europe in Dürer's days.

Second, in non-Western countries as well, linear perspective has been recognized as the most faithful representation of reality. For example, David Topper[38] cites a Chinese source rejecting the Western pictorial tradition as only a skill, yet admitting the faithfulness of linear perspective paintings. He also quotes a Japanese source praising linear perspective for its faithfulness.

Third, we should be skeptical about the ethnographic data purporting to show that some non-Western people cannot recognize photographs (which use linear perspective). R.K. Jones and Hagen[39] point out that Hudson's famous experiment failed to use correct linear perspective information. Rather, it used "intuitive approximations of such information" (*ibid.*, p. 209). That is, the Bantu and Western subjects were tested with the aid of *cartoon-like conventions* rather than with pictures with a proper linear perspective.[40] M.L. Leach, who used geometrically accurate material, found that none of the Shona children participating in his experiment failed to identify his pictures.[41] We may conclude that the results of Hudson's experiment and its replications are based on inadequate material.

---

[36] J.J. Koenderink et al. supported this idea with experimental results—Koenderink, A.J. van Doorn, and A.M.L. Kappers, "Depth Relief," *Perception*, XXIV (1995): 115–26, and "Pictorial Surface Attitude and Local Depth Comparisons," *Perception and Psychophysics*, LVIII, 2 (1996): 163–73.

[37] "Pictorial Depth Perception in Sub-cultural Groups in Africa," *Journal of Social Psychology*, LII (1960): 183–208.

[38] "Perspective on Perspective: Gombrich and His Critics," in R. Woodfield, ed., *Gombrich on Art and Psychology* (Manchester: University Press, 1996), pp. 78–100.

Fourth, the conventionalist conclusions of the ethnographic stories become even more unlikely when we note that 7-month-old infants and chimps can recognize depth and heights in (linear perspective) photographs.[42] Must we assume that infants have already been inculcated by their Western parents and the chimps by their Western zoos?

To conclude, there are serious doubts about the validity of the alleged ethnographic data, and there is a wealth of empirical data supporting the view that people who have not been inculcated with linear perspective representations experience these pictures as being convincing and faithful. Therefore, we need not worry about the Ultimate Conventionalist Argument. It is refuted by the empirical facts, and so is the conventionalist position: in spite of a nonlinear perspective tradition, people experienced linear perspective pictures as faithful.[43]

X. GOODMAN'S PRACTICAL CRITIQUE

Remember that Goodman's Practical Critique undid the Traditional Window Argument and it clashed with Hyman's analysis in terms of relative occlusion sizes as well. Does it not just as incontestably refute my analysis?

---

[39] "A Perspective on Cross-cultural Picture Perception," in Hagen, ed., Volume II, pp. 193–226.

[40] Perhaps this explains why a small but nonnegligible minority of Western subjects also failed to see the intended depth in Hudson's pictures.

[41] "The Effect of Training on the Pictorial Depth Perception of Shona Children," *Journal of Cross-Cultural Psychology*, VI (1975): 457–70. It is also relevant here that none of the Western subjects had problems in this experiment.

[42] Cf. R. Olson, A. Yonas, and R. Cooper, "Development of Pictorial Perception," in Hagen, ed., Volume II, pp. 155–91.

[43] Note that I do not argue here that linear perspective is an objective constraint for a faithful picture *because* both people from nonlinear perspective traditions and people from a linear perspective tradition experience linear perspective pictures as convincing. Though the universal reaction is very suggestive in this respect, and though this universal reaction does undermine the specific Goodmanian conventionalist claim on inculcation, the argument for objectivity is based on the explanation for this universal reaction, namely, the constancy scaling of the human perceptual apparatus. Please note as well that I do not deny that we may have to learn to read some linear perspective pictures.

After all, the photograph illustrating Goodman's example displays linear perspective and yet it does not look faithful or convincing. The hand thrust out on the photograph (figure 1) does not look like a real hand when it is thrust out.[44]

Note, however, that my analysis does not imply that every picture using linear perspective will look convincing or faithful. On my proposed analysis, a picture will do so only in those cases in which the constancy scaling within pictorial space corresponds strongly (enough) with constancy scaling within the real visual space. And we may now give another interpretation to Goodman's counterexamples: these are the very cases in which the equal constancy scaling assumption fails.

It is not that, in the photograph of the hand sticking out, the relative occlusion sizes are incorrect. They correspond to the occlusion sizes in the real world. The forward thrusting hand in real life does indeed occlude much of our field of vision. In this respect, the photograph of the hand is correct. However, in everyday life, size constancy usually takes over (our binocular viewing has no trouble with visual space close by) so that we do not notice how large the chunk is that is occluded, while apparently size constancy within that pictorial space is not strong enough to neutralize the large difference between the occlusion size of the hand and the occlusion size of the head.

The consequence is that the photograph does not look the way we see the hand in real life. So the photograph, though correct about relative occlusion sizes, is not faithful or convincing. What the "distorted" photograph brings to our notice is that, from a close station point, constancy scaling within the picture does not work nearly as efficiently as constancy scaling within real space. (There is a much stronger regression to the occlusion sizes). Thus, we get a conflict between what we see in real space and what we see in pictorial space.

This analysis is not just self-pleasing. There are independent reasons to expect that, with a close station point, the constancy scaling in pictorial space will be in trouble, while it still operates effectively in real, visual space. First, if there are objects in the foreground, as in the hand photograph, a close station point (as there is on the photograph) together with linear perspective will produce large contrasts in occlusion sizes in that very foreground. (The occlusion size of the hand on the photo is almost twice the occlusion size of the head on the photo.) This requires a large compensation from the constancy scaling. Second, for constancy scaling in pictorial space, the demand on our visual system is too great as both our binocular viewing of the picture and our view of the frame

---

[44] And indeed the mountains on the other photograph (used by Goodman) look much smaller to the observer than the real mountains. Here I shall not discuss the reason why.

(the border) give information about the flatness of the picture. The reason why the distortion problems do not arise in real, visual space is that these factors, which undermine pictorial space, do not undermine visual space; in fact they rather strengthen it. We view our real hands with both eyes from close quarters, but it is actually at this very close range that binocular disparity is an exceptionally strong depth cue. There is no frame, so there is no distraction due to a frame, and there is a clear foreground from our eyes to our hands. Our hands are then placed within a strongly articulated visual space. The great contrasts in occlusion sizes can be handled by size constancy.

Painters have long known about this problem and have had to find some means to conceal this conflict. Their practical—and objective—solution is to avoid a station point that is too close by. Leonardo DaVinci long ago warned against such a station point. Painters, then, are well aware of the distortions to which linear perspective can lead. They counteract these distortions by bringing in other objective constraints; objective in the sense that these constraints have their effect due to the way our perceptual system works, and not due to inculcation of some tradition.

In short, on my analysis, Goodman's counterexamples do not undermine the nonconventional status of linear perspective; they only point to limitations in its successful use.

XI. THE REMAINING WORRY

Given, however, that our binocular perception is optimally suited to real-life conditions while our binocular view diminishes the depth experience of a painting (by giving cues to the flatness of the canvas), and given that we see the frame (another factor which diminishes our pictorial depth perception), there must always be some discrepancy between constancy scaling in real, visual space and in pictorial space—even if linear perspective convergence is a powerful cause of pictorial depth and even if painters are aware of the limitations of linear perspective. So how can linear perspective succeed, since (in binocular viewing conditions) constancy scaling cannot be as effective within pictorial space as within visual space?

The answer is that we should recognize that our perceptions, both of real-space objects and of pictorial objects, are flexible. The amount of constancy scaling (both in real-life situations and in pictorial space) is not always the same. Depending on our attitude and training, our experience (both of real space and of pictorial space) may be more or less influenced by actual occlusion sizes. Often constancy scaling will dominate, but in many situations the pressure of the relative occlusion will be felt, that is, seen. I. Rock has emphasized this in his theory of the *double stance*: in our perceptions, the two different stances are always present.[45]

---

[45] "In Defense of Unconscious Inference," in Epstein, ed., pp. 321–73.

Here I do not, and need not, embrace Rock's theory of the double stance. We should, however, accept some flexibility of our perceptions. For instance, it is an established fact that painters have less constancy scaling than the average mortal. Practice teaches them actually to see the occlusion sizes of things. For normal mortals as well, size constancy scaling also fluctuates. Once we try (or manage) to see the two lonely monsters (figure 3) as running through a snow landscape, their sizes start to differentiate again. There are personal differences here, just as there are great personal differences in depth perception, both in real space and in pictorial space.[46] In other situations we can (to some extent) counteract the working of constancy scaling by concentrating on the relevant occlusion sizes.[47] It is tempting to point to this dual pull of perception, bringing along an ever-present conflict, to account for the less than complete constancy in many cases.

Note that these different degrees of regression to occlusion sizes (both in real, visual space and in pictorial space) lead to a great flexibility in our experience. This makes it a lot easier for a painting to look like reality.

Further, even if the object-in-pictorial-space looks too small or too large for the object depicted (because the constancy scaling in pictorial space would not increase the object enough in our experience), it will often be the case that there are smaller or larger objects in the world which do resemble the pictorial object as we experience it. In such a case the picture would not be faithful (it does not resemble the object depicted), but it would be convincing (because it resembles some imaginary world).

It is this flexibility in experience and diversity in the world which gives sufficient slack to cover the many situations in which the constancy scaling in real (visual) space and in pictorial space are somewhat different. With flexible appearances and diversity in the world, linear perspective paintings may well fall within the range of convincing pictures, in spite of the discrepancy between the constancy scalings.

These considerations explain to some degree how the inevitable difference in constancy scaling in real space and in pictorial space does not undermine the faithfulness or convincingness of paintings and pictures in linear perspective. I conclude that the Remaining Worry may be laid to rest.

XII. THINGS PUT TOGETHER

With the empirical data safely in place and Goodman's practical critique dealt with, we can put things together. We have seen that linear perspective in its convergence aspect does contribute strongly to the experience of pictorial depth. But once the painter has managed to obtain pictorial depth and pictorial space, constancy scaling starts to kick in within that pictorial space. So the painter must reduce the size of the painted objects. Here the other aspect of linear perspective comes in handy: the reduction of sizes according to converging lines, which comes to using relative occlusion sizes (they follow natural optics and hence correspond to linear perspective).

---

[46] See the articles by Koenderink et al.

[47] That is, by concentrating one can actually see the sizes of objects (more) corresponding to their relative occlusion sizes. In private conversation, Koenderink mentioned another example of flexibility in depth perception: often we begin to see more depth in a surrounding scene by concentrating on depth.

There was no empirical (scientific) support for the Constancy Argument in its strong version. The Equal Constancy Assumption could not be taken for granted. Our everyday data (for example, the picture of the two monsters in a corridor (figure 2)), however, confirm the Weak Constancy Argument. We have seen how this argument rules out reverse perspective and parallel perspective as candidates. We also have seen that the Eyewitness Principle excluded reduced perspectives. This leaves us with linear perspective as the only candidate.[48]

But how could it be a successful candidate when it is dependent on equal constancy scaling, while it seems reasonable to surmise that the constancy scaling within pictorial space must be weaker than constancy scaling within real space, as pictorial space has fewer depth cues? The quality of pictorial space must be less than the corresponding real, visual space. The Depth Argument gives part of the answer to this problem: linear perspective in its convergence aspect is on its own already a powerful creator of pictorial space. Yet binocular viewing conditions and the frame being in full view negatively affect our depth perception of pictorial space. However, the flexibility of our experience and the diversity in the world leave open the possibility of a faithful picture, in spite of these complications.

XIII. AN ANSWER TO GOODMAN'S THEORETICAL CRITIQUE

I am now in a position to answer Goodman's theoretical critique. First, note that—ironically—my argument shares Goodman's insight that a painter would be wise to take into account the different conditions in which the real object and the painting are seen. Indeed, painting is "not a matter of copying but of conveying." In my analysis, the painter has to counteract the constancy scaling which emerges as soon as he succeeds in creating some pictorial space. Thus, pictorial space brings along its dangers. The painter has to be aware that constancy scaling operates in different ways in real (visual) space and in pictorial space. For instance, the painter is in trouble if he chooses his station point too close by.

---

[48] Again, someone might want an explicit mention of curvilinear perspective at this place. In footnote 32, I argued, first, that it fails for reasons related to those which undermined (full-blooded) reduced perspectives and, second, that it is at least practically irrelevant.

Second, I am in no way tempted to deny that we need systems of representation to understand pictures and painting, and that we have different options here. But from all this it does not follow that linear perspective is a convention. The realist's point is that if we are interested in faithful or convincing pictures, the system of representation of linear perspective is the one which we should use for objective reasons, namely reasons which are dependent on how our perceptual system operates. Viewed in this light, linear perspective is just one of the objective means to handle properly differences in observational conditions. It is the system of representation that, when handled with care (and great skill), offers the possibility of a faithful and convincing picture.

XIV. CONCLUSION

There is indeed a convention in the Western world to use linear perspective in the attempt to create a faithful or convincing representation. But that is all the conventionality there is to linear perspective. Linear perspective is not conventional in the sense that conventionalists and postmodernists have in mind, namely that, given another tradition or given another education, we would have experienced pictures using another system of representation as (equally) faithful. Linear perspective is one of the main factors that evoke a faithful or convincing picture in people. In its aspect of convergence to a vanishing point it creates a convincing pictorial depth. In its other aspect, its size reduction aspect related to the relative occlusion sizes, it compensates the constancy scaling that affects our perceptions both in the real world and in the pictorial space.

We have also found that the real distortions in some linear perspective pictures are not an argument in favor of the conventionality of linear perspective. In fact, they demonstrate that, in addition to linear perspective, there are further nonconventional means that are needed for faithful and convincing pictorial appearances.

My defense of linear perspective as an objective, realist device to organize three-dimensional pictorial space does not, of course, rule out that there are other, radically different canons of representation and that we can understand (read) pictures using these different representational canons. However, the issue is whether these other canons yield equally faithful pictures of pictorial space, compared with linear perspective does. And the answer, for which I have argued, is that they do not.

ANTHONY A. DERKSEN

Radboud University Nijmegen

BOOK REVIEWS

Action in Perception. ALVA NOË. Cambridge: Massachusetts Institute of Technology Press, 2004. ix + 277 p. Cloth $38.00.*

This is a charming and ambitious book that combines careful attention to the phenomenology of experience with an appreciation of the psychology and neuroscience of perception. In some of its aims—for example, to show problems with a rigid version of a view of visual perception as an "inverse optics" process of constructing a static 3-D representation from static 2-D information on the retina—it succeeds admirably. As Nöe points out, vision is a process that depends on interactions between the perceiver and the environment and involves contributions from sensory systems other than the eye. He is at pains to note that vision is not passive. His analogy with touch is to the point: touch involves skillful probing and movement, and so does vision, although less obviously and in my view less centrally so. This much is certainly widely accepted among vision scientists—although mainstream vision scientists (represented, for example, by Stephen Palmer's excellent textbook[1]) view these points as best seen within a version of the inverse optics view that takes inputs as nonstatic and as including motor instructions (for example, involving eye movements and head movements).[2] The kind of point that Noë raises is viewed as important at the margins, but as not disturbing the main lines of the picture of vision that descends—with many changes—from the pioneering work of David Marr in the 1980s (and before him, from Helmholtz). But Noë shows little interest in mainstream vision science, focusing on nonmainstream ideas in the science of perception, specifically ideas from the anti-representational psychologist J.J. Gibson, and also drawing on Wittgenstein and the phenomenological tradition. There is a sense throughout the book of revolution, of upsetting the applecart. This is a review from the point of view of the applecart.

---

* I am grateful to Tyler Burge, John Campbell, Jakob Hohwy, Sean Kelly, Alva Noë, Christopher Peacocke, and Susanna Siegel for comments on a earlier draft.

[1] *Vision Science: From Photons to Phenomenology* (Cambridge: MIT, 1999).

[2] See, for example, Palmer's treatment of position constancy (objects do not appear to jump when the eye moves), on pp. 339–43, and optic flow, pp. 507–09. Helmholtz showed that the information that the brain uses to distinguish a sudden jerky motion of an object in the environment from a sudden jerky motion of the eye does not come from "afferent" (that is, input) sensors in the muscles that move the eye but rather from "efferent" motor commands to the eye. His experiment was simple: use the finger to move the eye, pushing from the side. The world appears to move, supporting the "efference copy" theory. As Palmer notes, an orthodox treatment can easily handle such inputs.

0022-362X/05/0205/259–72

My comments are in two parts, one mainly a priori, the other largely empirical: first, I will consider Noë's version of externalism in the light of the distinction between causation and constitution. Second, I will argue that on the most obvious reading of Noë's view, one that identifies perceptual experience with the skilled bodily exercise of "sensorimotor knowledge" (I will leave off the scare quotes in what follows) that includes visually guided action, there are empirical results that suggest that the exercise of such know-how does not reflect the phenomenology of conscious vision.

I. CAUSATION AND CONSTITUTION

Tyler Burge, Saul Kripke, and Hilary Putnam have argued for externalism about meaning and content: to take a familiar flamboyant version of it, two people with molecularly identical brains could nonetheless have different 'water'-thoughts because of differences in their physical and/or social environments. Fred Dretske, William Lycan, and Michael Tye hold that the phenomenal character of an experience is or at least supervenes on the experience's representational content, and so they use the Burge-Kripke-Putnam sort of externalism to support externalism about experience. They are Representationists about phenomenal character in the sense that they hold that the phenomenal character of an experience is exhausted by its representational content. Noë argues for what is in one way at least a more radical form of externalism about experience. His externalism is *vehicle* externalism rather than content externalism. The vehicles of contents are the physical items that have or express the contents—sentences for example. His analogy is to the view of Andy Clark and David Chalmers[3] that memory and calculation constitutively include props such as a diary or a pencil and paper. Similarly, according to Noë's view, the skilled active body partially constitutes the vehicle of experience.

Noë sometimes tries to frame the debate in a way that has his side arguing for a mere possibility, for example, "I have been arguing that, for at least some experiences, the physical substrate of the experience *may* cross boundaries, implicating neural, bodily, and environmental features" (221). However, these specks of caution float on a sea of exuberant declarations, such as "A neuroscience of perceptual consciousness must be an enactive neuroscience—that is, a neuroscience of embodied activity, rather than a neuroscience of brain activity" (227).

---

[3] "The Extended Mind," *Analysis*, LVIII (1998): 7–19. Clark and Chalmers do not apply their vehicle externalism to experience.

The leading idea of the book—what Noë calls the *enactive* view[4]—is a constitutive claim about experience: "Perceptual experience, according to the enactive approach, is an activity of exploring the environment drawing on knowledge of sensorimotor dependencies and thought" (228). What are sensorimotor dependencies (or contingencies) and what is knowledge of them? First, the 'sensori' in 'sensorimotor' is not supposed to be taken to be itself mentalistic. Noë is wary of the pitfall for his view of tacitly appealing to the content of experience in explaining the sensorimotor knowledge that is supposed to serve to explain the content of experience, so wary that he appears to adopt not one but two pieces of machinery for avoiding it. He says that strictly speaking sensorimotor knowledge—knowledge of sensorimotor contingencies—is knowledge of "the way sensory *stimulation* varies as you move" (78, emphasis added). The sensory side of the sensori-motor division is also spelled out in terms of what he calls "appearances," which are supposed to be totally objective. For example, the shape appearance of an object is given by any one of its projections on a plane that is orthogonal to the line of sight, for example, a projection on an imaginary window interposed between the eye and the object. Thus, the objective shape appearance of a round plate seen at an angle is the elliptical projection on such a plane, and so the plate objectively "looks" elliptical (from a certain angle). Sensorimotor knowledge is know-how concerning how objective appearances change as you move and the things you see move. The model for this know-how is knowing how to dance or gesture. The upshot is that perceptual experience, according to the enactive view, is the practical bodily exercise of sensorimotor know-how (see especially §3.5).[5] Noë's controversial conclusion is that perceptual experience does not constitutively supervene on the brain alone but only on the active body that is required for the skilled exercise of sensorimotor knowledge.

---

[4] This view has been pursued in other works by Noë and his collaborators, notably in Susan Hurley's *Consciousness in Action* (Cambridge: Harvard, 1998); and in J.K. O'Regan and Noë, "A Sensorimotor Approach to Vision and Visual Consciousness," *Behavioral and Brain Sciences*, XXIV, 5 (2001): 939–73. See also Francisco J. Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind* (Cambridge: MIT, 1991); and Dan Dennett's somewhat skeptical review of this book in *American Journal of Psychology*, CVI (1993): 121–26. An early precursor is H.L. Dreyfus, "Why Computers Must Have Bodies in Order to Be Intelligent," *Review of Metaphysics*, XXI (September 1967): 13–32.

[5] Christopher Peacocke noted (in conversation) that the devices I just mentioned do not avoid circularity. If you read about how to remove an appendix, that (plausibly) does not affect what it is like to see people who have appendixes, but it does add to one's knowledge of sensorimotor contingencies. The only way to rule out this booklearning kind of sensorimotor knowledge as determining the phenomenology of perception is to say that it is not implied by the perception itself. But that brings in the sensorimotor knowledge in explaining the perceptual content rather than the other way round.

Although I have described Noë as holding a constitutive externalist view of perceptual experience, I should say that there is considerable variation in the statement of the enactive view. Sometimes it is stated as a constitutive view about perception rather than perceptual experience: "perceiving is constituted by the exercise of a range of sensorimotor skills."[6] In addition to identifying perceptual experience with the exercise of sensorimotor skills, that is, skilled bodily exercise of sensorimotor knowledge, Noë also says that perceptual experience "draws on," is "constrained by," and is "enabled" by sensorimotor knowledge. We are often given declarations about the nature of perception or perceptual experience, but the terms change, frustratingly, from one statement to another. There are even parts of the book in which perceptual experience appears to be identified with a set of expectations rather than bodily activity. I believe that my interpretation of the enactive view as the claim that perceptual experience is the exercise of sensorimotor know-how is justified by the fact that it appears throughout the book, that it is often expressed emphatically ("Perceptual experience *just is* a mode of skillful exploration of the world" (194)) that it is emphasized in the last chapter (labeled "Brain and Mind: A Conclusion") which is on the mind-body problem and that this radical claim will inevitably be a source of controversy.

Noë supposes that the enactive view has a number of philosophical advantages. For example, he argues that it goes some way towards closing the famous "explanatory gap": Why is it that the neural basis of this experience is the neural basis of an experience like this instead of like something else or like nothing? In my view, the appearance of help here comes from a tacit conceptual functionalism (or even behaviorism). To put the point crudely, if you think the concept of a pain is the concept of a state that makes you say "Ouch" and the concept of an itch is the concept of a state that makes you scratch, it is a lot easier to close the explanatory gap for pain and itching and the difference between them.

Another line of argument that Noë offers in favor of the enactive view is based on impressive evidence that in many disparate kinds of cases, perceptual experience depends on sensorimotor contingencies. One point is emphasized by Brian O'Shaughnessy and Michael Martin[7]: our tactile experience of a solid object depends on sensing its resistance when we push against it, thus showing a causal dependence of experience on action. Another line of thought is experimental: in a variety of paradigms, changing sensorimotor contingencies changes the brain and experience itself. As Richard Held and Alan Hein showed in the 1960s, kittens that are active in exploring the environment have normal visual systems as compared with kittens that have the same stimulation but are passive.[8] Wearing inverting goggles results in considerable reorganization of perception. Hooking up a television camera to a pixel array grasped in a blind person's mouth allows for navigation through space, probably by recruiting "visual" areas of the brain.[9] These results are impressive but what they show is that sensorimotor contingencies have an *effect* on experience, not that experience is even partially *constituted* by—or supervenes, constitutively on—bodily activity. (To say that, for example, the moral facts supervene on the physical facts is to say that there can be no moral difference without a physical difference.)

---

[6] However, even in this case, the context (for example, the sentences before and after) make it clear that the topic is perceptual experience

[7] O'Shaughnessy, *The Will*, Volume 1 (New York: Cambridge, 1980), and *Consciousness and the World* (New York: Oxford, 2000); Martin, "Sight and Touch," in Tim Crane, ed., *The Contents of Experience: Essays on Perception* (New York: Cambridge, 1992), pp. 196–215. See also A.D. Smith, *The Problem of Perception* (Cambridge: Harvard, 2002).

This distinction between the claim that sensorimotor contingencies affect experience and the claim that experience is constituted by the exercise of sensorimotor know-how poses a major problem for many of Noë's arguments. The problem can be seen in stark form in Noë's discussion of dreaming. He says:

> Let us take it as settled that when we dream there is no dynamic exchange with the environment (although this might turn out not to be true), and let us accept that, therefore, neural states alone *are* sufficient for dreaming (*although this does not follow—e.g., the affective content of dream states may depend on endocrine gland activity, as waking emotional states do...*) (213, emphasis added).

Noë's thought here is that dream experience supervenes on the brain but perceptual experience does not. The suggestion raises the question of whether dream experience could really have a constitutive metaphysical basis that is different from that of perceptual experience, but I will pass over this issue, confining myself to commenting on what for Noë's purpose is just a parenthetical remark, a revealing aside. Noë supposes that if the affective content of dream states *depends causally on endocrine activity*, then the constitutive supervenience base for dream experience is not just the brain but has to include the endocrine system. But endocrine activity can affect dream experience without being part of the constitutive supervenience base for dream experience if endocrine activity causally affects that supervenience base. Indeed, the upshot of evidence on the brain basis of experience is that effects of the endocrine system on experience are mediated by effects on the brain itself and therefore do not challenge the orthodoxy that says that the brain is the minimal constitutive supervenience base for experience. (This is part of the applecart I mentioned earlier.)[10]

---

[8] "Movement-produced Stimulation in the Development of Visually Guided Behavior," *Journal of Comparative and Physiological Psychology*, LVI, 5 (1963): 872–76.

[9] See Hurley and Noë, "Neural Plasticity and Consciousness," *Biology and Philosophy*, XVIII (2003): 131–68; my "Spatial Perception via Tactile Sensation," *Trends in Cognitive Science*, VII, 7 (July 2003): 285–86; and Hurley and Noë, "Neural Plasticity and Consciousness: A Reply to Block, *Trends in Cognitive Sciences*, VII, 8 (August 2003): 342. Noë thinks that the subjects who have worn the goggles the longest end up finding the world to look as it did before they donned the goggles. Although I disagree with that gloss on the experiments, there can be no doubt that there is considerable experiential adaptation.

Let us be clear about what the issue is. The issue of the constitutive supervenience base for experience is the issue of what is—and is not—a *metaphysically necessary part of a metaphysically sufficient condition* of perceptual experience. That is, it is the issue of what is—and is not—part of the *minimal* metaphysically sufficient condition for perceptual experience (the minimal supervenience base). Noë's enactive view says that the skilled active body is part of that minimal supervenience base, whereas the view, which I hold and which I have labeled the orthodox view, is that nothing outside the brain is part of it.[11]

Importantly, the minimal supervenience base for an experience that occurs at time *t* is an instantiation of a physical property at *t*—according to the orthodox view. The Representationists mentioned earlier—Dretske, Lycan, and Tye—hold that the minimal supervenience base includes features of the environment from the past—for example, aspects of the evolutionary history of the individual and perhaps aspects of the individual's own lifetime of interacting with the environment. But the Representationists would agree with the orthodox view that if we hold these environmental variables fixed, only the features of the brain *now* are needed to determine the phenomenal character of experience *now*. And this is where the enactive view is more radical: the enactive view holds that even if we hold the evolutionary and immediate past environments fixed, there is still something outside the brain to be fixed in order to determine the phenomenal character of experience now, namely the activity of the body.

---

[10] Note that I am assuming only that the dreamer is experiencing, not that dream experience is perceptual experience.

[11] See my "Two Neural Correlates of Consciousness," *Trends in Cognitive Sciences*, IX, 2 (February 2005): 46–52; and Christof Koch, *The Quest for Consciousness* (Englewood, CO: Roberts, 2004). The orthodox view is also rejected by the Representationists mentioned earlier and by disjunctivists such as John McDowell and Mike Martin. See Susanna Siegel, "Indiscriminability and the Phenomenal," *Philosophical Studies*, CXX (2004): 90–112, for an excellent discussion of one version of disjunctivism.

Of course, there is often a process of perceiving that *involves* bodily activity—one moves closer or to the side to get a different view—but that should not be conflated with the very different idea that perceiving *is* an activity or that perceptual experience is an activity. And even if perceptual experience *depends* causally or counterfactually on movement or another form of activity, it does not follow that perceptual experience *constitutively* involves movement.

To illustrate the point, one might ask the question: suppose it was arranged so that the passive kitten in Held's and Hein's experiment had exactly the same brain goings on as the active kitten. The upshot of Noë's view is that that passive kitten still will not have the same experience as the active kitten.

Noë supposes that the main argument for drawing the line at the brain is the assumption that every experience that we can have can be produced by brain stimulation. He says:

> We are now able to produce very simple visual sensations such as the illusion of the presence of flashes of light ("phosphenes") by means of direct neural stimulation...from the fact that it is possible to produce *some* experiences, it does not follow that it is possible to produce *all* experiences. To assume, without further discussion, that we will someday be able to produce all perceptual experiences by direct neural stimulation, or that it is in principle possible to do so, is to assume too much. Indeed, it is to come close to assuming internalism about experience (211).

But the issue is not whether neural stimulation can *produce* every experience but whether if the relevant brain state were to come about—*somehow*—the experience would be instantiated. To suppose that the issue is one of how experiences can be produced is to shift the topic from a constitutive issue to a causal issue. Certainly, the causal sources of our experience include sensorimotor causal loops, but that does not settle the constitutive question.

Sometimes, Noë talks as if the issue is whether, when an experience is produced in a given environment, that environment is causally necessary for the experience.

> It may be that the only nomically possible world in which such a temporal series of brain states could occur would be one in which the animal were dynamically interacting with the very same kind of environment! To imagine the duplication of brain states is thus tacitly to appeal to the more extended setting in which those brain states are placed. Experience

> doesn't supervene on neural states alone, then, but only on neural states plus environmental conditions (223).

There is no plausibility in the claim that there is some nomologically possible series of brain states that can *only be produced by skilled interaction with a certain environment.* To see this point, we need only distinguish between the unlikely and the impossible. It is of course unlikely that all the air molecules in the room will all rush to one side, leaving you in a vacuum but it is not nomologically impossible. In fact, as I understand it, if the room constitutes a "quasi-ergodic" system (which rules out, for example, that the air molecules are perfectly elastic spheres which are hitting the perfectly reflective walls at 90°), it is guaranteed to happen "eventually"—according even to the Newtonian statistical mechanics of particles (although the sun may go out first). Once one adds quantum mechanics to the mix, the conclusion is even stronger. Any state of the brain that can be brought about by normal perception could perhaps occur—although with very low probability—by chance fluctuation.[12]

I will end this section with a side issue. I believe that interaction with the environment in evolutionary history and the history of the individual is constitutively necessary for an experience to have the representational content that it has. A brain in a vat (that has always been in the vat) does not have the same representational contents that that brain—even if relevantly internally the same—would have had in a normal perceptual environment. Indeed, a certain evolutionary background would be required for the brain in the vat to have anything that could qualify as a *perceptual* representation at all. (This is familiar content externalism.[13]) However, it is possible to resist the slide from externalism about the representational content of perception to externalism about the nature of the phenomenal character of perception.[14] In any case, as I mentioned earlier what Noë's book is about is not the issue of content externalism, but rather the quite different issue of vehicle externalism.

---

[12] When I pressed Noë on the constitutive/causal issue at the Workshop on the Fundamental Issues in Cognitive Science in Taipei, Taiwan (January 6–9, 2005), and at the Mind and Language Seminar at New York University (March 22, 2005), he disparaged the causal/constitutive distinction. But it should be noted that the book repeatedly makes what must be read as constitutive claims, including, sometimes, using the word 'constitutive', as in: "Most recent work on the relation of perception and action stops short of making the constitutive claim that defines the enactive standpoint: It does not treat perception as a kind of action or skillful activity (or as drawing on a kind of sensorimotor knowledge)" (18).

[13] See Burge, "Individualism and Psychology," *Philosophical Review*, xcv (1986): 3–45, "Individuation and Causation in Psychology," *Pacific Philosophical Quarterly*, LXX (1989): 303–22, and "Vision and Intentional Content," in Ernest Lepore and Robert Van Gulick, eds., *John Searle and His Critics* (Cambridge: Blackwell, 1991), pp. 195–213.

[14] See my "Mental Paint," in Martin Hahn and Bjørn Ramberg, eds., *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge: MIT, 2003), pp. 165–200, and Tyler Burge, "Qualia and Intentional Content: Reply to Block," in the same volume, pp. 405–15.

To sum up so far: sensorimotor know-how and perceptual experience are causally related, but that is no reason to think that they are constitutively related. I now turn to a very different line of probing having to do with a type of empirical finding that creates difficulty for the enactive view.

II. THE EMPIRICAL ISSUE

Thus far, I have not engaged much with the slippery issue of exactly what the enactive view really is. I will approach this issue by mentioning a line of empirical work that refutes the enactive view, at least on the most straightforward reading of it. Humans and other primates have two distinct visual systems, a conscious visual system that starts in the back of the head, moving to the bottom and side (the "ventral" system) and a much less clearly conscious "dorsal" system that goes from the back towards the top of the brain.[15] These two systems have very different properties. The ventral (conscious) system is slow, is oriented toward long-term visual planning of motion, and uses object-centered coding (objects are represented from a stereotypical point of view instead of the point of view that reflects the perceiver's current position). The dorsal system is fast, uses "egocentric" representations that are distance and orientation sensitive, has virtually no memory or color vision, and is used for the on-line visual guidance of action—for example, guiding the dribbling of a basketball down the court, avoiding obstacles. The last two sentences may sound like mumbo-jumbo to those who are unfamiliar with the concepts involved, so let me mention a few illustrations. Some of us have had the experience of running barefoot on the beach, our feet avoiding stones that we do not seem to see. The dorsal system—which feeds much more strongly than the ventral system to peripheral vision—is responsible for visually guided bodily movements that operate largely out of awareness.[16] There are a number of visual illusions that fool conscious vision but do not fool reaching and grasping motions that are guided by the dorsal system. (And some illusions that work in the reverse way!) It is the ventral system that is mainly relevant to perceptual experience, but—in a surprise result that rivals philosophers' science fiction scenarios—it is the dorsal system that guides action from moment to moment!

---

[15] See A. David Milner and Melvyn A. Goodale, *The Visual Brain in Action* (New York: Oxford, 1995), and Milner, *Sight Unseen* (New York: Oxford, 2004). An account aimed at philosophers is Andy Clark, "Visual Experience and Motor Action: Are The Bonds Too Tight?" *Philosophical Review*, CX, 4 (October 2001): 495–519.

[16] I. Schindler, N. Rice, R. D. McIntosh, Y. Rossetti, and Milner, "Automatic Avoidance of Obstacles Is a Dorsal Stream Function: Evidence from Optic Ataxia," *Nature Neuroscience*, VII (2004): 779–84; McIntosh, K.I. McClements, Schindler, T.P.Cassidy, D. Birchall, and Milner, "Avoidance of Obstacles in the Absence of Visual Awareness," *Proceedings of the Royal Society of London B*, CCLXXI (2004): 15–20.

Noë summarizes some of this research (describing it as "striking" (18)), but he does not grapple with what seems to me to be the real problem that this research poses for his perspective. (His response is in terms of teleology: perceptual experience is not *for* acting.) I can approach this problem by describing a dorsal/ventral difference which presupposes none of the theory that I have sketched. To understand it, you need to know that the fovea is the central area of the retina that is very densely packed with receptors. If you look at your thumb held at arms length, the fovea will be mainly occupied by the projection of your thumb. Here is the point: as James Danckert and Melvyn A. Goodale say: "The ventral 'perception' pathway...deals primarily with foveal vision—a fact that is consistent with the crucial role this pathway plays in the perception and recognition of objects and scenes.... In contrast, the dorsal 'action' stream, which...is known to play a critical role in the control of actions, such as goal-directed limb and eye movements, has a full representation of the visual field."[17] Goodale and Kelly Murphy[18] provide a dramatic example of this point. They presented 5 rectangular blocks to subjects at various positions in the visual field ranging from 5° to 70° off the line of sight. They compared accuracy of perceptual discrimination of one block from another with accuracy of grip via a device that measured the aperture between thumb and forefinger as subjects reached out to pick up one of the blocks. The basic finding is that subjects' grip accuracy is roughly the same at 5° as at 70°, whereas conscious perceptual discrimination is vastly worse (roughly one tenth as good by a standard measure) at 70° than at 5°.

You can get a first person appreciation of this point easily by holding up an object at 70° off of your line of sight. Features—including size—that you can see when you are looking straight at something are invisible or barely visible at an eccentricity of 70°—that is, they are inaccessible to conscious vision. But now try picking up something at 70°: you can do it pretty well. The upshot is that if the activity guided by sensorimotor knowledge with which the enactive approach identifies perceptual experience includes visually guided action, it *simply does not reflect the*

---

[17] "Ups and Downs in the Visual Control of Action," in Scott H. Johnson-Frey, ed., *Taking Action: Cognitive Neuroscience Perspectives on the Problem of Intentional Movement* (Cambridge: MIT 2003), pp. 29–64.

[18] "Action and Perception in the Visual Periphery," in P. Their and H.-O. Karnath, eds., *Parietal Lobe Contributions to Orientation in 3 D Space* (New York: Springer, 1997), pp. 447–61.

phenomenology of conscious vision. That is, sensorimotor knowledge—to the extent that it involves the visual guidance of action—is not true to perceptual experience.

At this point, however, we encounter a deep obscurity in the enactive view. Does the enactive view include visually guided action in the "activity" which it identifies with visual experience? The enactive view, as I have been understanding it, and as is stated often in the book, is that perceptual experience "is an *activity* of exploring the environment" drawing on sensorimotor know-how. Does the enactive view say that this activity (of exercising sensorimotor know-how) includes visually guided action? It is hard to see how skilled activity of exploring the environment drawing on sensorimotor knowledge-how could *fail* to involve visually guided action or what would be left of that skilled activity if visually guided action were removed.

However, there is one place in the book in which Noë explicitly denies that the sensorimotor know-how he is talking about includes visually guided action (see 90, where he describes visual guidance of action as "humdrum"). And there are a number of places in which he seems to take a view at odds with the often stated view that perceptual experience is an activity. Optic ataxia is a syndrome in which the dorsal system is damaged. If the ventral system is intact, the subject is impaired in visually guided action but apparently unimpaired in visual perception. Noë says that the optic ataxic

> does not undercut the enactive view, for from the fact that a patient suffers optic ataxia, it doesn't follow that he or she lacks the relevant sensorimotor knowledge. What would undercut the enactive approach would be the existence of perception in the absence of the bodily skills and sensorimotor knowledge, which, on the enactive view, are constitutive of the ability to perceive (12).

Here he seems to be identifying perception or perceptual experience with the mere possession of bodily skills and sensorimotor knowledge rather than their exercise in perceptually guided activity. Since the possession of sensorimotor know-how persists while experience changes, this cannot be quite what is meant, but perhaps there is some way other than visually guided action in which sensorimotor knowledge and bodily skills can be exercised and that might be supposed to be constitutive of visual experience?

One possibility is that the view is meant to be dispositional. But *dispositions* to visually guided action do not reflect the phenomenology of conscious vision any better than visually guided action itself. Dispositions to visually guided action are just as "dorsal" as visually guided action itself. Further, it is not clear that the optic ataxic *has* the relevant dispositions. Dispositional functionalists and behaviorists can and do deal with spinal cord paralysis by claiming that the paralysis affects the *manifestation* of the dispositions without affecting the dispositions themselves. For example, Behaviorists and dispositional functionalists sometimes handle paralysis by appeal to counterfactuals about how a person would move if his body were normal in certain ways ("paradigmatic embodiment," to use Sydney Shoemaker's term[19]), but such an appeal is only appropriate when the paralysis is caused by damage that affects the manifestation of the experience rather than the experience itself, for example, spinal cord damage rather than brain damage. But optic ataxia is a result of *brain* damage—not spinal cord damage—that affects the dispositions themselves, not just their manifestation. So dispositions will not save the enactive view.

There is a deflationary understanding of the enactive view that appears at some points in the book: that a perceptual experience is an instantiation of a *set of expectations* of how appearances will shift with movement. Some comments: first, the instantiation of a set of expectations need not involve any bodily skills or even a body, and so such a view would give up much of what is distinctive and revolutionary about the enactive view. (Recall: "Perceptual experience *just is* a mode of skillful exploration of the world.") Second, the set of expectations might be instantiated in the form of sensorimotor *mental imagery*, but this possibility provides no comfort for the enactive view since sensorimotor mental imagery could be entirely internal and not involve bodily skills in any way that requires a body at the time the motor imagery is exercised. What the enactive view would come down to is the rather un-applecart-upsetting view that perception involves a kind of mental imagery. Third, the point made in the first part of this review applies: even if expectations have an effect on perceptual experience, that is far from showing that perception is partially or totally constituted by the instantiation of expectations. Fourth, presumably the dorsal system involves at least implicit expectations for how objects in the environment shift with visual guided movements. So even the deflationary version of the enactive view under discussion would still clash with the facts about the two visual systems, since the enactive view would dictate that the (in fact unconscious) dorsal states are conscious. Fifth, it is unclear what the evidence in favor of such a view would be. Certainly, the points Noë makes about how changes in sensorimotor contingencies affect experience provide at best a weak argument for such a view.

If there is a constitutive role for anything sensorimotor in perception, I think it is likely to be a matter of one's spatial sense—a sense that is shared by many perceptual systems, including vision and proprioception, and that seems to be embodied in the dorsal system. One of the most interesting parts of the book is Noë's discussion of presence in absence, the phenomenology of the back of an apple that one does not see but senses nonetheless, or the cat that one sees—all of it—moving behind a picket fence. I think that presence in absence may be a matter of multimodal or amodal spatial imagery, and that that imagery may in part be motor imagery—since its brain basis appears to overlap with motor guidance systems in the dorsal visual system. Noë would have been on much stronger ground if he had restricted his enactive account to this aspect of experience instead of trying to capture all of experience. Sean Kelly has made a related suggestion: that knowledge of sensorimotor contingencies could provide an account of the background of perception rather than the foreground.[20] Here is a fact that supports both my view and Kelly's somewhat different view: subjects who have damage to the dorsal visual system that as I said is primarily responsible for visually guided action are worse at avoiding obstacles than in aiming for the main target.[21]

---

[19] "Embodiment and Behavior," in Amélie Rorty, ed., *The Identities of Persons* (Berkeley: California UP, 1976), pp. 109–37.

In an epilogue in the last four pages of the book, Noë seems to take back the claim that he has emphasized and spelled out in great detail throughout the book: that sensorimotor contingencies can be characterized objectively, suggesting that the enactive view must appeal to a primitive notion of life. "For living beings are already, by dint of being alive, *potentially conscious*" (230). The idea seems to be that what is common to all experience cannot be explained by appeal to knowledge of sensorimotor contingencies, but the difference between different perceptual contents can be so explained. "You give us a spark of consciousness, we'll give you the world" (230). I would suggest something more like the reverse: something sensorimotor has a better chance of explaining the spatial sense that is common to all or most perceptual experience than explaining what distinguishes different experiences.

The applecart that I have been defending against Noë's attack has two main tenets. First, the minimal constitutive supervenience base for perceptual experience is the brain and does not include the rest of the body. Second, although motor outputs and motor output instructions affect perceptual experience (as has been known since Helmholtz) much of perceptual experience can be understood in abstraction from such causes.

---

[20] At the NYU Mind and Language Seminar (March 9, 2004).

[21] See Schindler et al.; see also G.W. Humphreys and M.G. Edwards, "Automatic Obstacle Avoidance and Parietal Cortex," *Nature Neuroscience*, VII (July 2004): 693. This paper introduces the Schindler paper.

Although I have been heavily critical of Noë's book, I did find it interesting and intriguing. Not coincidentally, I think there is a grain of truth in the enactive view: perceptual experience is causally affected by sensorimotor contingencies and our sense of "presence in absence" may be a matter of sensorimotor imagery.

NED BLOCK

New York University

CAMBRIDGE

OUTSTANDING SCHOLARSHIP

The Changing Role of the Embryo in Evolutionary Thought
Roots of Evo-Devo
Ron Amundson
Cambridge Studies in Philosophy and Biology
$75 00: Hardback· 0-521-80699-2: 296pp

Aquinas: Disputed Questions on the Virtues
Edited by E.M. Atkins and Thomas Williams
Cambridge Texts in the History of Philosophy
$75.00· Hardback: 0-521-77225-7. c.340pp
$26.99: Paperback: 0-521-77661-9

Autonomy and the Challenges to Liberalism
New Essays
Edited by John Christman and Joel Anderson
$85 00: Hardback· 0-521-83951-3. 400pp

Foucault and Classical Antiquity
Power, Ethics and Knowledge
Wolfgang Detel
Modern European Philosophy
$75 00: Hardback: 0-521-83381-7: 292pp

John Locke's Politics of Moral Consensus
Greg Forster
$75.00: Hardback· 0-521-84218-2: 336pp

The Torture Papers: The Road to Abu Ghraib
Edited by Karen J. Greenberg and
Joshua L. Dratel
$50 00· Hardback: 0-521-85324-9· 1284pp

Kant and the Ethics of Humility
A Story of Dependence, Corruption and Virtue
Jeanine Grenberg
$75.00: Hardback 0-521-84681-1: 289pp

Kant: Notes and Fragments
Edited by Paul Guyer
The Cambridge Edition of the Works of
Immanuel Kant in Translation
$140.00. Hardback: 0-521-55248-6: 696pp

The Cambridge Companion to Montaigne
Edited by Ullrich Langer
Cambridge Companions to Philosophy
$65.00: Hardback: 0-521-81953-9: c.288pp
$22 99: Paperback: 0-521-52556-X

Naturalistic Hermeneutics
C. Mantzavinos
$65.00 Hardback. 0-521-84812-1: 200pp

Reconceiving Pregnancy and Childcare
Ethics, Experience, and Reproductive Labor
Amy Mullin
Cambridge Studies in Philosophy and Public Policy
$70.00. Hardback· 0-521-84438-X· 224pp
$24.99: Paperback: 0-521-60586-5

The Persistence of Subjectivity
On the Kantian Aftermath
Robert Pippin
$75.00: Hardback. 0-521-84858-X: 368pp
$28.99· Paperback· 0-521-61304-3

The Cambridge Companion to Maimonides
Edited by Kenneth Seeskin
Cambridge Companions to Philosophy
$70.00. Hardback: 0-521-81974-1· 408pp
$27 99: Paperback. 0-521-52578-0

Quality of Life and Human Difference
Genetic Testing, Health-Care and Disability
Edited by David Wasserman,
Jerome Bickenbach, and
Robert Wachbroit
Cambridge Studies in Philosophy and Public Policy
$70 00. Hardback: 0-521-83201-2: 288pp
$24 99. Paperback: 0-521-53971-4

Kant's Transcendental Proof of Realism
Kenneth Westphal
$80 00: Hardback: 0-521-83373-6: 310pp

Prices subject to change.

800-872-7423 • www.cambridge.org

CAMBRIDGE
UNIVERSITY PRESS

You can now subscribe to the

THE JOURNAL OF
PHILOSOPHY

using

PayPal

New Rates January 2005

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new
option, click on the PayPal icon at
www.jphil.org

Princeton University

The University Center for Human Values

LSR Visiting Fellowships

www.princeton.edu/values

The University Center invites applications from all disciplines for Laurance S. Rockefeller Visiting Fellowships. These fellowships will be awarded for the academic year 2006-07 to outstanding scholars and teachers interested in devoting a year in residence at Princeton writing about ethics and human values. Applicants typically have a doctorate or a professional post-graduate degree and cannot be in the process of writing a dissertation.

Fellows normally receive stipends of up to one-half their academic year salaries (not exceeding a maximum stipend set each fall) for the fellowship period, which extends from September 1st to June 30th. The fellows' home institutions are expected to provide at least half of their salaries in addition to all benefits.

How to apply:

For fellowships beginning in September 2006, applicants must submit a complete application by Tuesday, November 1, 2005. The detailed application procedure is available online at: www.princeton.edu/values.

Recipients of the Laurance S. Rockefeller Visiting Fellowships for 2006-07 will be announced by February 1, 2006.

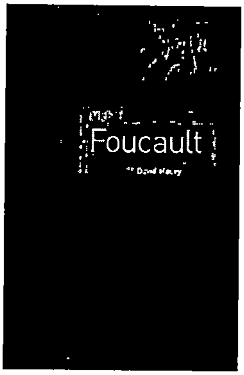
All application materials must be mailed to:

LSR Visiting Fellowships
University Center for Human Values
5 Ivy Lane
Princeton University
Princeton, New Jersey 08544

Princeton University is an equal employment opportunity, affirmative action employer. For more information about applying to Princeton and how to self-identify, please link to http://web.princeton.edu/sites/dof/ApplicantsInfo.htm.

new from

Reaktion Books



MICHEL FOUCAULT

David Macey

Though reticent about his personal life for most of his career, Michel Foucault, in the last years of his life, changed his stance on the relationship between the personal and the intellectual. David Macey demonstrates that these contradictions make it possible to relate Foucault's work to his life in an original and exciting way. He paints a portrait of Foucault in which the thinker emerges as a brilliant strategist, one who—while fiercely promoting himself as a maverick—aligned himself with particular intellectual camps at precisely the right moments.

Critical Lives

Paper $16.95

JEAN GENET

Stephen Barber

With an Introduction by Edmund White

This concise biography of Jean Genet and his work cuts directly to the intersection of thought and life that was essential to Genet's creativity. Arguing that Genet's life was an extraordinary spectacle in which the themes of his most revolutionary works were played out, Stephen Barber gives both the work and its singular inspiration in Genet's life their full due.

Critical Lives

Paper $16.95

ON GARBAGE

John Scanlan

On Garbage is the first book to examine the detritus of Western culture in full range—not only material waste and ruin, but also residual or "broken" knowledge and the lingering remainders of cultural thought systems. Scanlan considers how Western philosophy, science, and technology attained mastery over nature through what can be seen as a prolonged act of cleansing, as scientists and philosophers weeded out incorrect, outmoded, or superseded knowledge.

Paper $27.00

A PHILOSOPHY OF BOREDOM

Lars Svendsen

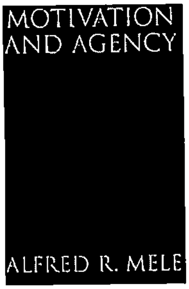
This witty exploration of the concept of boredom brings together observations from philosophy, literature, psychology, theology, and popular culture. Lars Svendsen examines boredom's pre-Romantic manifestations in medieval torpor, philosophical musings on boredom from Pascal to Nietzsche, and modern explorations into alienation and transgression by twentieth-century artists from Beckett to Warhol.

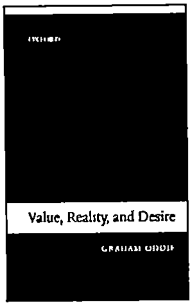
Paper $24.95

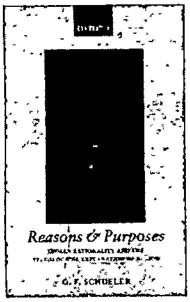
Distributed by

THE UNIVERSITY OF CHICAGO PRESS
1427 East 60th Street, Chicago, IL 60637
www.press.uchicago.edu

New from
Oxford







Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

New in Paperback!
Motivation and Agency
ALFRED R. MELE
What place does motivation have in the lives of intelligent agents? Mele's answer is sensitive to the concerns of philosophers of mind and moral philosophers and informed by empirical work. He offers a distinctive, comprehensive, attractive view of human agency.
paper May 2005 $19.95

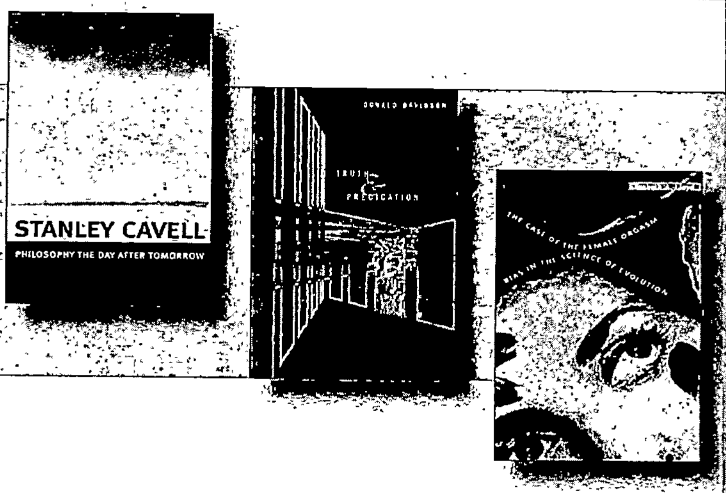
Epistemic Luck
DUNCAN PRITCHARD
Epistemic Luck is the first book to offer a rigorous philosophical examination of the concept of luck and its relationship to knowledge. In particular, Duncan Pritchard shows how a more nuanced understanding of the relationship between knowledge and luck can enable us to see our way past some of the most intractable disputes in the contemporary theory of knowledge.
May 2005 $55 00

Value, Reality, and Desire
GRAHAM ODDIE
What is the nature of value and how can we have knowledge of it? Is something valuable because we desire it or do we desire it because it is valuable? Graham Oddie argues that there are mind-independent irreducible facts about value, which are causally efficacious, and that we have knowledge of value by experiential acquaintance.
May 2005 $55.00

New in Paperback!
Reasons and Purposes
Human Rationality and the Teleological Explanation of Action
G. F. SCHUELER
What is it to act for a reason? Schueler rejects the dominant view that the beliefs and desires that constitute our reasons for acting simply *cause* us to act as we do, and argues instead for a view centered on practical deliberation—our ability to evaluate the reasons we accept. Schueler's account of "reasons explanations" emphasizes the relation between reasons and purposes, and the fact that the reasons for an action are not always good reasons.
paper May 2005 $24.95

New in Paperback!
Virtue Ethics
A Pluralistic View
CHRISTINE SWANTON
Christine Swanton offers a new, comprehensive theory of virtue ethics that addresses the major concerns of modern ethical theory from a character-based perspective. The book departs in significant ways from classical virtue ethics and neo-Aristotelianism, employing insights from Nietzsche and other sources.
paper May 2005 $29.95

JKC 030205


STANLEY CAVELL
PHILOSOPHY THE DAY AFTER TOMORROW
DONALD DAVIDSON
TRUTH & PREDICATION
THE CASE OF THE FEMALE ORGASM
BIAS IN THE SCIENCE OF EVOLUTION

PHILOSOPHY THE DAY AFTER TOMORROW

STANLEY CAVELL

"One of our most imaginative philosophers, Cavell can always be counted upon to provoke his readers to join him as he soars to dizzying new philosophical heights."

—HENRY L. CARRIGAN JR., *LIBRARY JOURNAL*

Belknap Press • new in cloth

TRUTH AND PREDICATION

DONALD DAVIDSON

"Davidson was a distinguished philosopher, and the first three chapters of this book constitute his principal statement about the concept of truth. I consider it one of the most important philosophical writings about truth of its time."

—CHARLES PARSONS, PROFESSOR OF PHILOSOPHY, HARVARD UNIVERSITY

Belknap Press • new in cloth

THE CASE OF THE FEMALE ORGASM

Bias in the Science of Evolution

ELISABETH A. LLOYD

"Lloyd's book isn't just a 'must-read,' it's a must-own, must-cite, and must-assign-to-one's-students. William James once defined philosophy as 'an unusually stubborn effort to think clearly.' This is the most stubborn effort I've ever seen to think clearly about female orgasm."

—RACHEL MAINES, AUTHOR OF *THE TECHNOLOGY OF ORGASM*

new in cloth

WWW.HUP.HARVARD.EDU HARVARD UNIVERSITY PRESS

ALIPORE CAMPUS
04/01/06

THE JOURNAL OF PHILOSOPHY

VOLUME CII, NUMBER 6
JUNE 2005

page

273 *The Simplicity of Identity* André Gallois
303 *The Subtraction Argument for Metaphysical Nihilism* David Efird and Tom Stoneham
326 NEW BOOKS: TRANSLATIONS

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Christopher Peacocke, Philip Pettit, Carol Rovane, Achille C. Varzi. *Consulting Editors*: David Albert, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2005

SUBSCRIPTIONS (12 issues)

Individuals	$45.00
Libraries and Institutions	$100.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2005 by the Journal of Philosophy, Inc. ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CII, NO. 6, JUNE 2005

THE SIMPLICITY OF IDENTITY*

David Lewis writes about the so-called *problem of trans-world identity*:

> More important, we should not suppose that we have here any problem about *identity*.
>
> We never have. Identity is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be.

He adds:

> We *do* state plenty of genuine problems in terms of identity. But we *needn't* state them so. Therefore they are not problems about identity.[1]

I will argue that this view of identity and the puzzles it putatively generates is mistaken. Identity is no simpler than any other relation. I plan to expose a feature of our thought and talk that generates the illusion that, as a relation, identity enjoys a distinctive simplicity.[2] One motivation for doing so is that until the illusory simplicity of identity is exposed, certain solutions to what are genuine puzzles about identity cannot get a fair hearing.

---

* I would like to thank audiences at the Australasian Philosophical Association Conference in Adelaide, the University of Rochester, and Syracuse University for helpful discussion. I would especially like to thank John Bigelow, David Chalmers, Mark Heller, Eric Hiddleston, Kevin Kukla, Kris McDaniel, Tom McKay, Ishani Maitra, Brendan Murday, Calvin Normore, Joshua Spencer, and Megan Wallace for challenging comments.

[1] *On the Plurality of Worlds* (New York: Blackwell, 1986), pp. 192–93.

[2] Here I may seem to be ignoring the tradition, represented by Wittgenstein in the *Tractatus*, which has it that identity is not a relation at all. I am not. *One* reason for thinking that identity is not a relation is the same as the one inspiring the thought that it is a simple relation.

0022-362X/05/0206/273–302

What are the puzzles, and what the solutions? One type of puzzle concerns fission and fusion cases such as amoebic division and the Ship of Theseus. Another is illustrated by the relation between a statue and the statue-shaped clay composing it. Solutions to these puzzles have included an appeal to both contingent and vague identities. Both of these solutions have been taken to be ruled out by arguments invoking a principle constitutive of identity, Leibniz's Law. The arguments fail, but showing why requires saying what produces the illusion that identity is simple.

In the first section, I discuss a convention governing our repeated use of referring expressions, and say how attending to some implications of that convention being in force helps to defuse the so-called modal argument for the necessity of identities. In the second I respond to objections to this way of meeting the modal argument. In the third I apply the results of the first to a parallel argument against indeterminate identities. Finally, in the last section, I sketch the application of the strategy used to undermine the arguments discussed in the first three to Frege's Puzzle.

I. NECESSARY IDENTITY

The claim that identity is contingent is thought to be ruled out by a well-known argument advanced by Saul Kripke and Ruth Barcan Marcus.[3] The claim that identity is vague is thought to be ruled out by an argument advanced by Nathan Salmon and Gareth Evans.[4] I begin with a discussion of the Kripke-Barcan Marcus argument. Here is one version of that argument. Suppose that 'a' and 'b' are rigid codesignators. Hence:

(1) $a=b$

But;

(2) $\Box(a=a)$

Employing λ abstraction to construct a predicate attributing the property of being necessarily identical with a, we supposedly have:

(3) $(\lambda x)(\Box x=a)a$

---

[3] For Kripke's statement of the modal argument, see "Identity and Necessity," in Milton Munitz, ed., *Identity and Individuation* (New York: University Press, 1971), pp. 135–64; for Marcus's, see *Modalities* (New York: Oxford 1993).

[4] Evans gives one version of the argument against indeterminate identities in "Can There Be Vague Objects?" *Analysis*, XXXVIII (1978): 208. Salmon gives a different one in *Reference and Essence* (Princeton: University Press, 1981).

Since, by Leibniz's Law, if (1) then a and b share all properties in common. Hence (1) and (3) yield:

(4) $(\lambda x)(\Box x=a)b$

which in turn yields:

(5) $\Box(a=b)$

Even those inclined to question (3) accept (2). There is no reason to accept (3), but there is also no reason to accept (2). (2) tends to be endorsed unhesitatingly because it is conflated with a different claim which is true. Moreover, the claim that (2) is conflated with is one that, unlike (2), does not imply (3).

Consider the following sentence containing two tokens of the type name 'Aristotle':

(S1) 'Aristotle was a philosopher, and Aristotle was married to Jackie Onassis'.

Suppose Jayne uses (S1) to make a statement. It is just conceivable that Jayne thinks the late husband of Jackie Onassis spent his time philosophizing when not running a shipping line. Alternatively, Jayne may subscribe to a bizarre hypothesis about Aristotle the ancient philosopher being reincarnated as a twentieth-century shipping magnate. Charity recommends a different hypothesis. When she utters (S1) Jayne is violating a convention I will call the *identity convention,* which is a default convention governing the use of different tokens of the same type name in the same linguistic context. Restricting attention to different tokens of the same type name showing up in the same sentence, the identity convention says this:

IC: Unless contextually or otherwise indicated, two tokens of the same name type occurring in the same sentence can be taken to have the same referent.

The identity convention is deeply, and properly, entrenched. A language that made provision for co-referring tokens of the same type term that was not governed by the identity convention would be unworkable. Nevertheless, it is the identity convention that produces the illusion that identity is a simpler relation than it is. Correlatively it also gives premise (2) of the above version of the modal argument, $\Box(a=a)$, an undeserved appearance of self-evidence. In order to show how the identity convention obscures the disputability of (2), I will employ the following strategy. I will compare the identity convention's effects with the effects of a parallel convention, entitled the *same height convention,* governing the relation of being the same height. Like identity, the relation of being the same height is an equivalence relation. Unlike identity, all would agree that the same height relation sometimes holds contingently. The point of introducing the same height convention is this. Consider a sentence *S* which states that a pair of individuals are the same height. If *S* is governed by the same height convention, we will find the following holds. *S* can only be used to state a truth. Moreover it can be known *a priori* that *S* states a truth. Finally, *S* appears uninformative in the same way that an identity sentence of the form '$a=a$' appears uninformative.

We will find that the identity convention bestows the same features of being comparatively uninformative and being knowable *a priori* to be truth expressing on the identity sentences it applies to. Moreover, it bestows those features for the reason that the same height convention bestows the features in question on a sentence attributing the same height relation. Being subject to the same height convention gives us no reason to suppose that a sentence attributing the same height relation states a necessary truth. Likewise, once the effects of being subject to the identity convention are taken into account, we will see that we have as little reason to suppose that a sentence governed by the identity convention expresses a necessary truth as we do for thinking that a true sentence governed by the same height convention expresses a necessary truth.

Consider the following sentences:

(S2) 'Jayne is the same height as Mary'.
(S3) 'Jayne likes Mary'.

If they express truths, (S2) and (S3) express contingent truths. Here is the same height convention.

> SH: Suppose a name *N* in a sentence is starred, and *a* is the referent of *N*. In that case, for *a* to be the referent of *N* it must be the same height as the referent of any name immediately preceding *N* in the same sentence.

The following illustrates the same height convention. Suppose we introduce the starred name 'Mary*'. Compare sentence (S3) with:

(S4) 'Jayne likes Mary*'.

Let us agree that, in order for (S4) to express a proposition, 'Mary*' must refer.[5] Here is a condition that must be satisfied if (S4) is to express a true proposition. Since, the name 'Jayne' immediately precedes the starred name 'Mary*' in (S4), the referent of 'Mary*' must be the same height as the referent of 'Jayne'.

---

[5] The agreement is only provisional. It would be incautious to assume that an indicative sentence expresses a proposition only if the (used rather than mentioned) referring expressions in it refer to something that exists. Nevertheless, I believe, the best way to demonstrate the misleading effects of the identity convention is to initially make that assumption, and then show how, without adversely impacting on the main argument, it can be dispensed with.

The relevant point to note about (S4) is this. First, (S4) expresses a contingent proposition. Moreover, without violating the same height convention, (S4) could have expressed a false proposition. Finally, (S4) is no less informative than (S3).

There is a further, and for our purposes important, point to note about (S4). I said that, for example, (S4) is at least as informative as (S3). Arguably (S4) is more informative than (S3). Suppose Sally reads (S4) knowing that the same height convention is in place, and believes it to express a truth. In addition, suppose, antecedently to reading (S4), Sally has no beliefs about the distinctive relations holding between the referents of 'Mary*' and 'Jayne'. If so, on reading (S4), Sally will acquire the following pair of beliefs about the referents of 'Mary*' and 'Jayne':

(i) Jayne likes Mary*

and:

(ii) Jayne is the same height as Mary*

(S4) is the sentence used to express (i). On inspecting (S4), Sally comes to believe (i). She does so because (S4) says that (i). (S4) says that Jayne likes Mary*. On inspecting sentence (S4) Sally also comes to believe (ii). She does *not* do so because (S4) says that (ii). (S4) does not say that Jayne is the same height as Mary*. (S4) does not express a proposition that entails anything about Jayne and Mary*'s relative heights.

If (i), the proposition expressed by (S4), does not entail (ii), why does Sally come to believe (ii) when she becomes acquainted with (S4)? She does so for pragmatic reasons of a familiar type. She knows that (i) is being conveyed to her by means of an utterance of (S4), and that (S4) contains the starred name 'Mary*'. She also knows that 'Mary*' is subject to the same height convention. Hence she is entitled to conclude that, if (S4) expresses a proposition and does not violate the same height convention, then Jayne is the same height as Mary*. Sally does believe that (S4) expresses a true proposition. That is why she infers (ii).

(S4) is the sentence 'Jayne likes Mary*'. When she comes to be acquainted with (S4), Sally learns both (i) and (ii). Sally learns (i) because (S4) states that (i). We may say that (S4) semantically attributes the liking relation. Sally, on being acquainted with (S4), does not learn (ii) because (S4), unlike sentence

(S5) 'Jayne is the same height as Mary*'.

semantically attributes the same height relation. She learns (ii) because she employs a metalinguistic premise together with her background knowledge of the same height convention to conclude that (ii) is true. We may say that (S4), though it does not attribute it, metalinguistically imputes the same height relation.

Here once again is the same height convention:

> SH: Suppose a name *N* in a sentence is starred, and *a* is the referent of *N*. In that case, for *a* to be the referent of *N* it must be the same height as the referent of any name immediately preceding *N* in the same sentence.

Since we are supposing SH is known to be in force, (S4) metalinguistically imputes an additional relation to the one semantically attributed by (S3). Hence, (S3) is less informative than (S4). Now consider the sentence expressing (ii):

(S5) 'Jayne is the same height as Mary*'.

(S5) is less informative than:

(S2) 'Jayne is the same height as Mary'.

Moreover, and this is the crucial point, the degree to which the informativeness of (S5) differs from (S2) is the same as the degree to which informativeness of:

(S6) 'Jayne is identical with Mary'.

differs from the informativeness of:

(S7) 'Jayne is identical with Jayne'.

(S5) is less informative than (S2) because the relation, the same height relation, that (S5) semantically attributes is the same as the one it metalinguistically imputes. Moreover, though the proposition expressed by (S5) is not *a priori* knowable, the following is:

(ii*) If 'Mary*' conforms to the same height convention, and 'Jayne' immediately precedes 'Mary*' in sentence (ii), then Jayne is the same height as Mary*.[6]

---

[6] We might think that it is not (ii*) that is *a priori* knowable but: if 'Mary*' conforms to the same height convention, and 'Jayne' immediately precedes 'Mary*' in sentence (ii), then the referent of 'Jayne' is the same height as the referent of 'Mary*'. The latter is *a priori* knowable, but so is (ii*) since it is *a priori* knowable that 'Jayne' refers to Jayne, and that 'Mary*' refers to Mary*.

On superficial inspection (ii*) is necessarily true. It is not. What is necessarily true is the following:

(ii#) If 'Mary*' conforms to the same height convention, 'Jayne' immediately precedes 'Mary*' in Sentence (ii), and, with the exception of the starred name it contains, sentence (ii) has its standard English meaning, then the referent, if any, of 'Jayne' is the same height as the referent, if any, of 'Mary*'.

Sentence (S5) is 'Jayne is the same height as Mary*'. (ii#) tells us that if (S5) expresses a proposition, it expresses a true proposition. Since (ii#) is necessary, it invites us to draw a distinction that has played a prominent role in recent philosophical discussion. It invites us to distinguish between a sentence expressing a necessary truth, and it being necessary that a given sentence expresses a truth. (S5), if it expresses anything, expresses a truth. That much is guaranteed by the relation semantically attributed by (S5) coinciding with the one it metalinguistically imputes. Nevertheless, if it expresses a truth, the truth expressed by (S5) is manifestly contingent.

How does this apply to the argument for the necessity of identities? Consider:

(S7) 'Jayne is identical with Jayne'.

The argument for the necessity of identities turns on sentences of the form of (S7) expressing necessary truths.[7] Here is what I called the identity convention:

IC: Unless contextually, or otherwise, indicated two tokens of the same name type in the same sentence can be taken to have the same referent.

In a given linguistic context, unless the identity convention is suspended as in the case of one token of 'Aristotle' referring to the philosopher and another in the same sentence referring to Jackie's ship owning husband, an utterance of (S7) metalinguistically imputes the relation of identity holding between the referent of the first occurrence of 'Jayne' and the referent of its second occurrence. As in the case of sentence:

(S5) 'Jayne is the same height as Mary*'.

---

[7] Of course for a reason irrelevant to assessing the modal argument we might agree that (S7) does not express something necessarily true. (S7) may express no necessary truth because 'Jayne' fails to refer. For ease of exposition, for now, I am ignoring this complication. Later we will consider the consequences of replacing (S7) with '$(\exists x)$ x=Jayne → Jayne=Jayne'.

the relation metalinguistically imputed by (S7) is the same as the relation semantically attributed by the same sentence.

Call the condition that the relation semantically attributed by a sentence is the same as the relation metalinguistically imputed the *coincidence condition.* The preceding discussion of (S5) yielded the following conclusions. If a sentence *S* satisfies the coincidence condition then:

(a) Even if the proposition expressed by *S* is contingent, it is necessary (with appropriate qualifications about, for example, keeping fixed its actual meaning) that *S* expresses a true proposition.
(b) It is knowable *a priori* (again with the same qualifications) that *S* expresses a true proposition.
(c) If *S* satisfies the coincidence condition, and sentence *S'* does not, but *S* and *S'* semantically attribute the same relation to exactly the same things, then, ceteris paribus, *S* is less informative than *S'*.

There is, I contend, no reason to suppose that:

(S7) 'Jayne is identical with Jayne'.

expresses a necessarily true proposition even if Jayne exists. Even if Jayne exists there is no reason to endorse:

(iii) □(Jayne is identical with Jayne)

(iii) may be true, but it is no more obviously so than, say: if Jayne is identical with Mary, then □(Jayne is identical with Mary). The illusion that (iii) is evidently true is, I am maintaining, produced by something, sentence (S7), satisfying the coincidence condition, that ensures the following is true:

(iii#) □(If (S7) has its ordinary English meaning, conforms to the identity convention, and 'Jayne' has a referent, then (S7) expresses a truth)

The coincidence condition is the condition that the relation semantically attributed by a sentence is the same as the relation it metalinguistically imputes. How can we test whether (S7) satisfying the coincidence condition is responsible for generating an illusion of necessity? We can do so by considering a case where the identity convention is explicitly suspended.

Consider a puzzle case in which, according to some, an identity contingently holds. At some time t_1 Tibbles, an entailed cat, has a proper part, Tib, which consists in all of Tibbles at t_1 except for Tibbles's tail. The only relevant change that takes place between t_1 and t_2 is that Tibbles has lost its tail. Suppose that at t_2:

Tib_1: Tibbles=Tib

If Tib_1 is true at t_2, there is reason to hold it to be contingently true. After all, Tib_1 looks false at t_1. At this point, the argument for not taking Tib_1 to be contingently true gets to work beginning with the premise:

Tib_2: $\Box$(Tibbles=Tibbles)

Tib_2 looks self evidently true. But suppose we introduce the following pair of stipulations about the referents of different occurrences of 'Tibbles' in sentence

Tib_3: 'Tibbles is identical with Tibbles'

The first exempts 'Tibbles' from the scope of the identity convention. We explicitly allow that the first token of 'Tibbles' in Tib_3 may, though it need not, refer to something different from the second token of the same name in that sentence. In addition, we stipulate the following. If it is not preceded by any occurrence of 'Tibbles' in a sentence, that token of 'Tibbles' refers to the cat complete with tail at t_1. If a token of 'Tibbles' is preceded by a token of the same name in the same sentence, the succeeding token of 'Tibbles' refers to Tib, a proper part of Tibbles, at t_1.

Now consider:

Tib_3: 'Tibbles is identical with Tibbles'

At t_2 if we adhere to the above stipulation, it becomes highly controversial whether Tib_3 expresses a true proposition. Recall that, in addition to suspending the application of the identity convention to Tib_3, we have made a stipulation with the following effect. The initial token of 'Tibbles' in Tib_3 refers to the cat with a tail at t_1. The second token of 'Tibbles' in Tib_3 refers to a cat like proper part of the cat with the tail at t_1. So is Tib_3 true at t_2? That depends on whether Tib is identical with Tibbles at t_2. Defenders of the contingent truth of Tib_1 will say that Tib_3 expresses a truth at t_2. Defenders of the necessity of identities will deny that Tib_3 expresses a truth at t_2. If Tib_3 does express a truth, it is nearly as controversial whether it is necessarily true. Hence invoking the above stipulations and, thereby, suspending the application of the identity convention to 'Tibbles', makes the necessary truth of what is expressed by Tib_3 exactly as debatable as the necessary truth of Tib_1.[8]

---

[8] Here I am confining myself to answering one major argument against the contingency of identity. It is a lengthy task, one falling well outside the scope of this paper, to say why the contingency of identity provides the best solution to venerable identity puzzles. I undertake that task in my *Occasions of Identity* (New York: Oxford, 1998).

In order to turn Tib_3 into something that manifestly expresses a truth, we need to reinstate the identity convention. Doing so, I have argued, transforms Tib_3 into a sentence that can, with the usual qualifications, be known *a priori*, to express a truth.[9] But doing so does not alter the modal status of the proposition expressed by Tib_3. If exempted from the scope of the identity convention, Tib_3 is not self-evident or knowable *a priori*, and may or may not express a necessary truth. If subject to the identity convention, Tib_3 is self-evident and knowable *a priori*, but, still, may or may not express a necessary truth.

II. OBJECTIONS

I claimed that an identity sentence *S* being subject to the identity convention in virtue of satisfying the coincidence condition has the following consequences:

(a) Even if the proposition expressed by *S* is contingent, it is necessary that *S* expresses a true proposition.
(b) It is knowable *a priori* that *S* expresses a true proposition.
(c) If *S* satisfies the coincidence condition, and sentence *S′* does not, but *S* and *S′* semantically attribute the same relation to exactly the same things, then, ceteris paribus, *S* is less informative than *S′*.

Why are (a)–(c) consequences of a sentence being subject to the identity convention, and, hence, satisfying the coincidence condition? For the moment, I will set aside (c). (c) is relevant to Frege's Puzzle, but has no particular relevance to the modal argument, or the argument against indeterminate identities. So I will reserve to the later discussion of Frege's Puzzle investigating why a sentence *S* being subject to the identity convention implies (c).

What of (a) and (b)? Why should we accept that (a) is a consequence of the identity convention applying to *S*? Call the defense of the contingency of $a=a$ which relies on (a) the *identity convention* (IC) *account*. The first objection to the IC account I want to consider argues there is no reason to accept (a) is a consequence of the identity convention applying to *S*. To see how it goes consider, again, the Tib-Tibbles case.

In the Tib-Tibbles case Tib is at one time, putatively, a proper part of Tibbles, but, says the defender of contingent identities, at another time identical with Tibbles. On that view we have:

---

[9] If Tib_1 is false, and Tibbles is distinct from Tib, reinstating the identity convention with respect to Tib_3 would have the following effect. It would ensure that 'Tibbles' refers to at most one of Tib and Tibbles.

Tib_1: Tib=Tibbles

together with:

Tib_1^*: $\Diamond \neg$ (Tib=Tibbles)

Someone who endorses the modal argument will deploy the following premise to refute the conjunction of Tib_1 with Tib_1^*:

Tib_2: $\Box$(Tibbles=Tibbles)

Consider the sentence:

Tib_3: 'Tibbles=Tibbles'

According to the IC account, the identity convention ensures that Tib_3 expresses a truth without that sentence necessarily expressing a necessary truth. This is because if Tib_3 is subject to the identity convention, it will satisfy the coincidence condition. The relation, in this case identity, semantically attributed by Tib_3 will be identical with the relation metalinguistically imputed by that sentence. Moreover, satisfying the coincidence condition is supposed to explain further relevant features of Tib_3 and like sentences. It is supposed to explain why Tib_3 appears to express something knowable *a priori*. It is also supposed to explain why Tib_3 appears less informative than Tib_1 even if the latter is true.

What is the effect of Tib_3 satisfying the coincidence condition? For our purposes, the main one is this. Anyone who knows that Tib_3 is subject to the identity convention will be able to infer from the repetition of the same type name in Tib_3 that the first token has a referent identical with the referent of the second. Combining that with the knowledge that 'Tib' refers to Tib yields the conclusion that Tib is identical with Tibbles. The same information can be gleaned from recognizing what truth condition attaches to Tib_3. So the operation of the identity convention does explain why Tib_3 will deliver the same information twice over. What, goes the objection, satisfying the coincidence condition fails to explain is why Tib_3 must express a truth. It also fails to explain why Tib_3 expresses something relatively uninformative and knowable *a priori*.

Here is one way to illustrate this objection. We could adopt the following convention: writing any sentence whatever in red ink conveys the information that snow is white. So, for example, writing the sentence 'Grass is green' would convey two items of information. It would convey, in virtue of saying it, that grass is green. It would also convey by being written in red ink, and being subject to the previously described convention, that snow is white. Call this convention the *red* *ink convention.* Suppose, following the red ink convention, we write the sentence 'Snow is white' in red ink. If it is subject to the red ink convention, writing that sentence in red ink will convey the information that snow is white twice over. Despite that, the sentence 'Snow is white' written in red ink, even if governed by the red ink convention, could fail to express a truth, gives no appearance of expressing something knowable *a priori*, and is no less informative than the same sentence written in blue ink. Clearly satisfying the coincidence condition is not enough for Tib_3 to have the features it needs to counter the relevant version of the modal argument.

Call this first objection to the IC account the *information duplication objection.* Briefly the information duplication objection goes as follows. What the IC account tells us is that an identity sentence of the form '$a=a$' conveys the same information twice over by both semantically attributing and metalinguistically imputing the identity relation. What that fails to explain is why such an identity sentence is uninformative, and can be known *a priori* to be necessarily truth expressing even if the truth it expresses is contingent.

The task is to explain why an identity sentence of the form '$a=a$' appearing to express something knowable *a priori* and necessary gives us no reason to think that it does so. From the point of view of explaining why that is so, the implication of the identity convention, that an identity sentence of the form '$a=a$' presents the same information twice over, supplies an essential part, but only a part, of that explanation. Equally crucial is the mechanism that leads an identity sentence of the relevant kind to metalinguistically impute the relation it semantically attributes.

I illustrated the duplication objection by means of the hypothetical red ink convention. The red ink convention stipulates that writing any sentence in red ink conveys the information that snow is white. If the red ink convention is in force, writing in red ink the sentence 'Roses are purple' conveys some misinformation about the color of roses together with some information about the color of snow. Moreover, writing 'Snow is white' in red ink conveys the same information twice over. Now suppose that snow had not been white. In that case, if the red ink convention had been in force, and 'Snow is white' had been written in red ink, that sentence would still have expressed a proposition albeit a false one.

In this respect, let us compare the red ink with the identity convention in light of the assumption, call it the *reference assumption,* that a sentence expresses no proposition unless its referring terms refer. Consider the identity sentence introduced earlier:

(S7) 'Jayne is identical with Jayne'.

The identity convention tells us that something is the referent of the second token of 'Jayne' in (S7) only if it is identical with the referent of the first token of that name in the same sentence. Given the reference assumption, could (S7) express a false proposition? If it did, the referent of the second token of 'Jayne' would be distinct from the referent of the first. If so, (S7) would either be exempt from the scope of the identity convention, the identity convention would fail to be in force, or be violated by (S7). We are assuming that the identity convention is in force. We are also assuming that there are no contextual factors making (S7) exempt from, or in violation of, the identity convention. The only remaining alternative is that (S7) expresses no proposition, and so no true proposition.

The combination of the identity convention with the reference assumption dictates the following. If an identity sentence such as (S7) expresses a proposition at all, it has to express a true one. That is why, if the identity convention and reference assumption are both in place, (S7), and like identity sentences, are guaranteed to express truths without there being any guarantee that the truths they express are necessary. That is also why (S7) can be known *a priori* to express a truth without the truth expressed having to be knowable *a priori*. More exactly, that is why it can be known *a priori* that, if an identity sentence of the relevant kind is subject to the identity convention, and the reference assumption is correct, then that identity sentence will express a truth.

The reference assumption says that if any of the expressions used with the intention of referring in an indicative sentence fail to refer that sentence fails to express a proposition. The reference assumption is contestable. If the IC account presupposes the truth, let alone *a priori* knowability, of the reference assumption, so much the worse for that account.

Can the IC account dispense with the reference assumption? We see that it can when we recall that a defender of the modal argument is not committed to, for example:

(S7) 'Jayne is identical with Jayne'.

expressing a necessary truth. All that a defender of the modal argument is committed to is that:

(S7*) '$(\exists x)(x{=}\text{Jayne}) \rightarrow \text{Jayne}{=}\text{Jayne}$'

expresses something necessarily true. Suppose the antecedent of (S7*) does not express a truth. In that case, the conditional expressed by

(S7*) will be true. Suppose the antecedent of (S7*) does express a truth. In that case, the two tokens of 'Jayne' in (S7*) will refer. The identity convention ensures that they refer to the same thing. So, the truth of what is expressed by (S7*'s) antecedent, together with the application of the identity convention guarantees that the conditional expressed by (S7*) is true without guaranteeing that it is necessarily so.

The second objection to the IC account can be dealt with more straightforwardly. It goes as follows. Consider the following uncontroversially necessary truth: 2 is a prime number. That the sentence '2 is a prime number' expresses that truth depends on that sentence being governed by contingent linguistic conventions. No one should be tempted to conclude from this that there is no reason to believe that necessarily 2 is prime. Why then should we conclude from the sentence 'Tibbles=Tibbles' being governed by the identity convention that we have no reason to believe that 'Tibbles=Tibbles' expresses a necessary truth?

We know that, given the linguistic conventions governing its use, the sentence '2 is prime' must express a truth. But we know that only because we know the truth expressed by '2 is prime' is necessary. Knowledge of the identity convention being in force enables us to know that 'Tibbles=Tibbles' must, if it expresses anything, express a truth *whether or not we know that the truth it expresses is necessary.*

Here is another way to put the point. We have knowledge of the conventions governing the use of the expressions in '2 is prime'. That knowledge enables us to understand what is said by that sentence. What knowledge of the relevant linguistic conventions will not, by itself, tell us is whether '2 is prime' expresses a truth. We need independent grounds that go beyond knowledge of linguistic conventions to tell that '2 is prime' expresses a truth.

In contrast, knowledge of the conventions governing the use of the expressions in 'Tib=Tib' suffices for knowing that sentence expresses, if anything, a truth.

The following three objections have a common theme. It is that the IC account gains whatever plausibility it has from exclusively focusing on a version of the modal argument involving repeated proper names. In my opening discussion, I presented a version of the modal argument involving the repeated use of a constant whose key premise is:

(2) $\Box(a=a)$

An alternative version of the modal argument replaces (2) with:

(2*) $\Box(\forall x)(x=x)$

(2*) is stated without using anything like a proper name or constant. It is stated using only a variable.[10] Hence, pointing out the consequences of the identity convention governing the use of proper names or constants does not speak to a version of the modal argument that is stated without using either.

There is an obvious reply. Repeated tokens of the same type variable are quite as much governed by an extension of the identity convention as proper names are governed by the identity convention as originally stated. Consider the open sentence:

OS: '$(x=x)$'

OS contains two tokens of the same type variable 'x'. Unless otherwise indicated we should take it that whatever is assigned to the second token is identical with whatever is assigned to the first. The identity convention applies to repeated tokens of the same variable quite as much as it applies to repeated tokens of the same name or constant.

By itself, this observation does not entirely meet the objection we are considering. Here is one way to reinstate it. Suppose we do not demand that repeated uses of the same variable obey the identity convention. In particular, let us suspend the presumption that the two tokens of 'x' in sentence (2*) conform to the identity convention. Even so (2*) is true. This is because, whether or not the identity convention dictates the same thing must be assigned to repeated tokens of 'x' in sentence (2*), the same thing must be assigned to those tokens because they are bound by a single quantifier.

This version of the objection gets things the wrong way round. We can bind the variable in sentence (2*) with a single quantifier only because it is understood that repeated tokens of that variable conform to the identity convention. Suspend the convention enforcing that understanding, and sentence (2*) simply becomes ill formed.

I began with a quote from David Lewis. Part of the passage quoted runs: "There is never any problem about what makes something identical to itself; nothing can ever fail to be." This suggests a different, albeit related, objection to the IC account. Suppose that we grant there is no reason to think that say:

Tib$_3$: 'Tibbles=Tibbles'

expresses a necessary truth even if the two occurrences of 'Tibbles' co-refer, and conform to the identity convention. Still we have to admit that each thing is necessarily identical with itself. So we should admit that Tibbles is necessarily identical with Tibbles even if the necessary self identity of Tibbles is not adequately expressed by Tib_3. Moreover, the necessary identity of each thing with itself is the crucial premise driving the modal argument, however that premise is best expressed.

---

[10] Kripke's own statement of the argument in "Identity and Necessity" utilizes only variables.

Of course what this overlooks is that the above points made about the identity convention apply, with one qualification, to the sentence:

(S8) 'Tibbles is identical with itself'.

In virtue of its meaning in English, 'itself' functions to back-refer in (S8) to the referent of the subject term 'Tibbles'. Which is to say that 'itself' in, for example:

(S9): 'Tibbles is licking itself'.

will take as its referent only something identical with the referent of 'Tibbles'. (S9) does not semantically attribute the identity relation. Rather (S9) metalinguistically imputes that relation. So the relation metalinguistically imputed by (S9) is distinct from the one it semantically attributes. In virtue of containing 'itself' (S9) metalinguistically imputes the same relation, identity, as (S8). The difference is that the relation metalinguistically imputed by (S8) is the same as the one it semantically attributes.

Tib_3 is the sentence 'Tibbles=Tibbles'. There is a difference between (S8) and Tib_3 that needs to be noted. I would say that it is not in virtue of its meaning that the second token of 'Tibbles' in Tib_3 should be taken to refer to something identical with the referent of the first.[11] The identity convention is defeasible, and does not determine the meaning of the expressions governed by it. In contrast 'itself' back-refers to Tibbles in (S8) because of its meaning.

What follows from this is that (S8) satisfies the coincidence condition in virtue of its meaning. Hence the meaning of (S8) guarantees that it expresses a truth. Still there is no guarantee that the truth expressed by (S8) is necessary.

We have examined objections to the IC account with the following feature in common. Each accuses the IC account of trading on using repeated tokens of the same name or constant to state the target premise of the modal argument. Here is a further objection of the same type.[12] Consider the sentence:

---

[11] I put it this way rather than saying: it is not in virtue of its meaning that the second token of 'Tibbles' in Tib_3 refers, as opposed to should be taken to refer, to something identical with the referent of the first. Suppose one thinks that token names have meanings, and that their meanings determine their referents. In that case one may well think that, since the two tokens of 'Tibbles' in Tib_3 refer to Tibbles in virtue of their meanings, it is in virtue of its meaning that the second refers to something identical with the referent of the first.

(S10) '*a*=*a*'

Suppose we introduce some new expressions. Instead of using '=' to express identity let us use '→'. So '*a*→*a*' says that *a* is identical with *a*. Now suppose we avoid repetition of the same constant in stating what is stated by '*a*→*a*' by adopting the following convention. We use the symbol '↩' to indicate that the referent of the only token of '*a*' in '*a* ↩' is identical with that thing. No repetition of the same type constant here.

Certainly we may dispense with using a term back referring to *a* in stating that *a* is identical with *a*. Having adopted the above convention governing the use of '↩' we can state the identity of *a* with *a* by means of '*a* ↩'. Still it remains to be asked how the relevant convention can be introduced. I introduced it by writing the following sentence:

(S11) 'We use the symbol' '↩' to indicate that the referent of the only token of '*a*' in '*a* ↩' is identical with that thing'.

(S11) contains the referring expression, 'that thing'. (S11) would fail to introduce the bent back arrow convention unless 'that thing' is taken to refer to something identical to what is referred by 'the referent of the only token of '*a*' in '*a* ↩''. So, (S11) will give the meaning of in '*a* ↩' only if (S11) is governed by the identity convention.

Of course there are alternative sentences we can use to explain the meaning of '*a* ↩'. We can explain it by using any of ''*a* ↩' means that *a* is identical with *a*', or ''*a* ↩' means that *a* is identical with itself' or 'the arrow doubling back means that *a* stands in the identity relation to that same thing'. Selecting any of these is to explain the meaning of '*a* ↩' by using a sentence that, to employ the terminology introduced in the previous section, both metalinguistically imputes and semantically attributes the identity relation; that is, by employing a sentence satisfying the coincidence condition.

A different version of essentially the same objection to the IC account appeals to rigidified descriptions. In the Tib-Tibbles case, suppose, instead of using a proper name to refer to Tibbles we use a rigidified description such as 'the only cat actually at place *p* at time

---

[12] I am indebted to Calvin Normore for this objection.

t'. It is in virtue of the meaning of the contained description rather than being subject to an identity convention contextually subject to suspension that the following is guaranteed to express a truth:

TIB: 'The only cat actually at *p* at *t*=the only cat actually at *p* at *t*'.

If that is right then, surely, TIB expresses a necessary truth. Moreover, if what TIB expresses is necessarily true, and Tibbles has the property of being the only cat actually at *p* at *t*, then Tibbles has the further property of being necessarily the only cat actually at *p* at *t*. That modal property will be shared by Tib if Tibbles is identical with Tib.

We need to ask whether TIB expresses a necessary truth because it has the form '*a*=*a*', or because of the property Tibbles has to have, being the only cat actually at *p* at *t*, in order for it to satisfy the description 'the only cat actually at *p* at *t*'. The answer is, of course, to satisfy the relevant description it must have the property of being the only cat actually at *p* at *t*.[13]

Here is a different type of objection to the IC account. The IC account would be in trouble if it implied that a sentence of the form '*a*=*a*' cannot express a necessary truth. The account has no such implication. It is neutral about which identities hold of necessity and which contingently.

The IC account applies a distinction familiar from so called two dimensional modal semantics. Because of an apparent dissimilarity between identity sentences and sentences to which the distinction in question seems to apply it may be questioned whether the distinction between a sentence having to express a truth and it expressing something that has to be true could apply to the former. In the case of a sentence like:

(S12) 'I am here'.

we need to distinguish between that sentence having to express a truth, and it expressing something that has to be true. Why do we need to draw this distinction in the case of (S12)? We need to do so because (S12) will express different propositions in different worlds. What this allows for is that (S12) will always express a true proposition with respect to any world even if it expresses a proposition in some world that is false in some other world.

---

[13] Does it follow from Tib being identical with Tibbles, and, so, sharing with Tibbles the necessary property of being the only cat actually at *p* at *t* that Tib is not contingently identical with Tibbles? It does not. All that follows is that Tib and Tibbles are necessarily *actually* identical. It does not follow that Tib and Tibbles are necessarily identical.

The relevant distinction between necessarily expressing a truth and expressing a necessary truth applies to (S12) because (S12) exhibits indexicality. Tib_3 does not appear to exhibit indexicality; still less does a sentence like 'Tibbles is identical with itself', or '$(\forall x)(x=x)$'.[14] How, in the case of these sentences, can we explain their having to express truths coming apart from their expressing necessary truths?

The task is to explain how a sentence such as Tib_3 can, in virtue of being subject to the identity convention, express different propositions with respect to different worlds. The explanation I favor presupposes a currently popular view of propositions: the so-called Russellean view.

Consider:

Tib_1: Tib=Tibbles

and the sentence:

Tib_3 'Tibbles=Tibbles'

Recall how we fixed the referents of the names in the sentence expressing Tib_1. 'Tibbles' refers to a cat that earlier has a tail, but is later without one. 'Tib' refers to cat that is earlier a proper part of Tibbles, and, after Tibbles loses its tail, coincides with Tibbles.

Suppose, that after Tibbles loses its tail, Tib_1 becomes true. Suppose that, after Tibbles loses it tail, Tib is identical with Tibbles. Now consider the proposition that Tibbles is hungry. According to the Russellean view I have in mind, the proposition that Tibbles is hungry has as its constituents the cat Tibbles, and the property of being hungry. Moreover, any proposition with exactly the same constituents will be identical with the proposition that Tibbles is hungry. In particular, given Tib_1, the proposition that Tib is hungry will, after Tibbles loses its tail, be identical with the proposition that Tibbles is hungry.

Now suppose that, even though Tib_1 is true, it is contingently so. Tib is contingently identical with Tibbles. Consider a possible world *W* in which Tib is not identical with Tibbles. For simplicity suppose Tib does not exist in *W*, but Tibbles does exist in that world; something that we are entitled to regard as coherent if Tib is only contingently identical with Tibbles. Since Tib is actually identical with Tibbles, in the actual world the proposition that Tib is hungry is identical with the proposition that Tibbles is hungry. Hence, on the view being considered, the sentence 'Tibbles is hungry' expresses at the actual world the proposition that Tib is hungry. What proposition does that sentence express at *W*? Since Tibbles exists in *W*, but Tib does not, and we are supposing that a proposition is individuated by its constituents, the sentence 'Tibbles is hungry' expresses the proposition Tib is hungry at the actual world without expressing that same proposition at *W*. Hence, the sentence 'Tibbles is hungry' expresses a proposition at one world that it does not express at another even though it expresses some proposition at each one.

---

[14] 'Tibbles is identical with itself' contains an indexical, but does not seem to exhibit indexicality in the sense that its truth value is context dependent.

Let us apply this to the proposition expressed by Tib$_3$. Since Tib$_1$ is true, both tokens of 'Tibbles' in Tib$_3$ refer in the actual world to both Tib and Tibbles. Hence Tib$_3$ expresses at the actual world the proposition that Tib is identical with Tibbles. As they are actually used, the two tokens of 'Tibbles' refer in *W* to Tibbles, without referring to Tib in that world. Hence Tib$_3$ does not express the proposition that Tib is identical to Tibbles at *W*. Moreover, Tib$_1$, the proposition that Tib$_3$ does express at the actual world, is false at *W*. It does not follow that Tib$_3$ expresses a false proposition at *W* because at that world it does not express Tib$_1$.

A final objection to the IC account goes like this. I sought to explain the way in which the identity convention misleads us into thinking that a sentence such as:

(S7) 'Jayne is identical with Jayne'.

manifestly expresses a necessary truth by initially identifying the like effects of the same height convention on a sentence such as:

(S5): 'Jayne is the same height as Mary*'.

The objection is this. Even with the same height convention understood to be in force, no one would be tempted to suppose that (S5) expresses a necessary truth. In contrast, many of the most able philosophers suppose (S7) to express something that holds of necessity. Why the difference?

In effect, the answer to this objection is contained in its statement. It is obvious that the same height relation only holds contingently. So, there is no danger that the operation of the same height convention will mislead anyone into thinking that (S5) can only express a necessary truth. Of course, no like obviousness attaches to the view that identities hold only contingently.

III. INDETERMINATE IDENTITY

Here is one version of the standard argument for there being no indeterminate identities. Let '*Ip*' mean: it is indeterminate that *p*, and '*Dp*' mean: it is determinate that *p*. We have:

(1′) $I(a=b)$

But:

(2′) $D(a=a)$

(2′) yields:

(3′) $(\lambda x)(D(x=a))a$

Suppose:

(4′) $a=b$

Applying Leibniz's Law to (3′) and (4′) gives:

(5′) $(\lambda x)(D(a=x))b$

from which it follows that:

(6′) $D(a=b)$

which contradicts (1′). Hence (1′) implies:

(7′) $\neg(a=b)$

If (7′) is true then it is not true that *a* is identical with *b*. So if (7′) is true, *a* is determinately distinct from *b*. (1′) cannot be true.

In the case of the modal argument, the counterpart premise to: (2′) is, of course:

(2) $\Box(a=a)$

I argued that, once the consequences of the identity convention being in force have been noted we lack any reason to endorse (2). Does the same hold for (2′)? Not quite. The identity convention does give (2′) a spurious appearance of self-evidence. Nevertheless, it is a slightly more complicated task to say why for the following reason.

My claim is that, though it is necessary that '$a=a$' (here insert the requisite qualifications) expresses a truth, there is no warrant for thinking that the truth expressed by '$a=a$' is necessary. That is, there is no warrant for thinking that:

(2i) $(\forall p)$('$a=a$' expresses $p \rightarrow \Box p$)

is true. Yet, (2) is true only if (2i) is. Likewise (2′) is true only if the proposition expressed by '$a=a$' is determinately true. That is, (2′) is true only if

(2d) $(\forall p)$('$a=a$' expresses $p \rightarrow Dp$)

In the case of:

(2) $\Box(a=a)$

the argument went like this. Even if (again with appropriate qualifications about that sentence having its standard meaning, and so on) it is necessary that '$a=a$' expresses a truth, recognition that the identity convention is in place deprives us of any reason to think that the truth expressed by '$a=a$' is necessary. It would be nice if a parallel claim about indeterminate identities could be given a straightforward defense.

The parallel claim about indeterminate identities goes like this. Even if the identity convention ensures that '$a=a$' determinately expresses a true proposition, the existence of that convention deprives us of any reason to think that the proposition expressed by '$a=a$' has a determinate truth value.

Of course, this last claim is not defensible. If it is determinate that '$a=a$' expresses a true proposition then the proposition expressed by '$a=a$' cannot lack a determinate truth value. Despite this, the existence of the identity convention does yield a reply to the Leibniz's Law argument against indeterminate identities.

Here, once again, is the identity convention:

> IC: Unless contextually, or otherwise, indicated, two tokens of the same name type in the same sentence can be taken to have the same referent.

It is natural to take 'the same' in the sentence expressing IC to mean: is determinately identical with. Suppose we do, and '$a=a$' is subject to the identity convention. If P is the identity proposition expressed by '$a=a$', then, as we have seen, the identity convention ensures that P is true. Hence, no identity proposition can both be expressed by '$a=a$' and lack a determinate truth value. Since it will not express a false proposition, we can be confident that if '$a=a$' expresses a proposition at all, it expresses a true one. So, either the sentence mentioned in (2i), '$a=a$', expresses no proposition, or it expresses one that is true.

At this point the defender of indeterminate identities has, at least, the following options. One is to admit that if '$a=a$' expresses a proposition then:

(2′) $D(a=a)$

is true. So the issue then becomes whether, given that '$a=a$' remains subject to the identity convention, it expresses a proposition. The conclusion of the argument against indeterminate identities:

(5′) $D(a=b)$

turns out to be exactly as disputable as the claim that (2′) expresses a proposition. A better way for the advocate of indeterminate identities to go is, I think, this. The identity convention is:

> IC: Unless contextually, or otherwise, indicated two tokens of the same name type in the same sentence can be taken to have the same referent.

IC is revocable. Suppose, in order to be no questions against indeterminate identities, we revoke it in the present case, and replace it with:

> DC: Unless contextually, or otherwise, indicated, two tokens of the same name type in the same sentence can be taken to not be determinately distinct.[15]

DC tells us that the second token of '*a*' in '*a*=*a*' refers to something that is not determinately distinct from the referent of the first token of '*a*' in that sentence. DC leaves it open whether the referent of the second token of '*a*' in '*a*=*a*' is indeterminately identical with the referent of the first token of '*a*'.

Replacing IC with DC avoids begging the question against the indeterminate identity theorist. Does it beg the question against her opponent who deploys the above argument against indeterminate identities? It does not. Replacing IC with DC would beg the question against the determinate identity theorist if DC ensured that '*a*=*a*' lacks a determinate truth value. Of course, it does no such thing. Even if '*a*=*a*' is governed by DC rather than IC, it may be that the referent of the first occurrence of '*a*' is determinately identical with the referent of its second.

IV. FREGE'S PUZZLE

Here is Frege's own version of the puzzle that has come to bear his name. We consider:

(1) Hesperus=Phosphorus
(2) Hesperus=Hesperus

Sentence (1) appears to differ from sentence (2) in the following ways. Sentence (1) is informative in a way that sentence (2) is not. In addition, sentence (2) appears to express something that can only be known *a priori*, but sentence (1) appears to express something that can only be known *a posteriori*. The puzzle of course is to explain how sentence (1) and sentence (2) can differ in these ways since both sentences state that the same planet, Venus, stands in the same relation, identity, to that very planet.

---

[15] Earlier I remarked that eliminating the identity convention altogether would make a language unworkable. It is now time to enter a qualification. The purposes served by IC are equally well served by DC.

According to an increasingly popular view, sentences (1) and (2) do not differ in at least one of the relevant ways. Call this the *no difference view.* On one popular version of the no difference view (1), like (2), can be known, *a priori,* and, in the sense in which the informativeness of a sentence is a function of its truth condition, sentence (1) is no more informative than (2). In addition, sentence (1) is *less* informative than it appears to be. Call this the *deflationist* version of the no difference view.[16] An alternative version of the no difference view goes like this. (2), like (1), is something that can only be known *a posteriori.* Moreover, sentence (1) is *no* less informative than it appears to be. Call this the *inflationist* version of the no difference view.[17]

I claim that noting the consequences of the identity convention being in force is relevant to solving Frege's Puzzle. It has at least the following relevance. Pointing to the effects of referring terms being governed by the identity convention helps to explain how either version of the no difference view could be true. In particular, it helps to explain how the less frequently defended inflationist version could be true.

Before embarking on that explanation, something needs to be said about why it may seem that the identity convention could not be relevant to solving Frege's Puzzle. The reason why it may appear irrelevant is that Frege's Puzzle does not seem to essentially be a puzzle about identity.[18] Consider:

(3) Hesperus is the same size as Phosphorus.

and:

(4) Hesperus is the same size as Hesperus.

Though they concern a relation that holds between identicals, sentences (3) and (4) do not express identities. Nevertheless, it seems, sentences (3) and (4) pose the same puzzle as sentences (1) and (2). Sentences (3) and (4) appear to have the same truth conditional content. Each says of a certain planet, Venus, that it stands in the relation to being the same size as to what is in fact that very same planet. Despite that, (4) looks *a priori* knowable, but (3) only *a posteriori* knowable. Also, sentence (4) seems less informative than sentence (3) to the same degree that sentence (2) is less informative than sentence (1).

---

[16] I call this the deflationist version for the following reason. According to this version, contrary to appearances, it takes no more than *a priori* reflection to know the proposition expressed by sentence (1).

[17] I call this second version of the no difference view inflationist because, again contrary to appearances, *a priori* reflection is not enough to have knowledge of the proposition expressed by (2).

[18] That it is not essentially a puzzle about identity has been convincingly argued by Nathan Salmon in his book *Frege's Puzzle* (Cambridge: MIT 1986). Plausible though it is, as we will see, Salmon's argument overlooks a way in which identity plays an essential role in generating the puzzle even in those cases where it does not seem to.

Sentences (3) and (4) do pose the same puzzle as the one posed by the identity sentences (1) and (2). Even so, the puzzle posed is, essentially, a puzzle about identity. Giving the defense of the no difference view which features the identity convention will, hopefully, make clear why that is so.

Consider, again, the version of Frege's Puzzle posed by:

(1) Hesperus=Phosphorus
(2) Hesperus=Hesperus

Earlier I listed three consequences of a sentence *S* being subject to the identity convention, and explained why the first two are consequences. It is now time to explain why the one particularly relevant to Frege's Puzzle is also a consequence. Here is (c):

> (c) If *S* satisfies the coincidence condition, and sentence *S′* does not satisfy that condition, but *S* and *S′* semantically attribute the same relation to exactly the same things, then, ceteris paribus, *S* is less informative than *S′*.

Sentence (2) is subject to the identity convention. Since sentence (2) expresses an identity it satisfies the coincidence condition. Sentence (1) semantically attributes, but does not metalinguistically impute the identity relation. So, unlike sentence (2), sentence (1) does not satisfy the coincidence condition. What needs explaining is why this difference between sentences (1) and (2) renders (2) less informative than (1).

Knowledge can be lost through the acquisition of further knowledge. I know that the person I have just seen is Jones. Jones's, unbeknown to me demented, mother tells me that it is not Jones, but her twin sister that I have just seen. I cease to know that I have just seen Jones. I lose that item of knowledge because I have acquired the knowledge that Jones's mother has just informed me that I have just seen Jones's twin sister.

In this case, I lose knowledge that I already have because I acquire new knowledge. It could just as easily have worked in reverse. I might have been blocked from acquiring the knowledge that I have just seen Jones by earlier knowing that Jones's mother has told me her tale.

Suppose we think of the truth conditional information potentially conveyed by a sentence as (encapsulated in) the proposition(s) it expresses. In addition, let us provisionally suppose that (1) is the proposition expressed by sentence (2). So, sentences (1) and (2) potentially convey exactly the same information. How then are we to explain the following? That, standardly, someone who understands sentence (2), and believes it states something true, would not come to know that (1)?

The explanation is provided by attending to the identity convention. Even if it is implicit, there is common knowledge of the identity convention. Now, consider someone making a statement by uttering sentence (2). Suppose, as would normally be the case, that the name in sentence (2) is known by the audience to be subject to the identity convention. What the audience knows, even in advance of sentence (2) being uttered, is that, if the speaker's linguistic intentions are fulfilled, the two tokens of 'Hesperus' that it will contain will refer to the *same* thing. Hence, the audience knows that if the first and second tokens of 'Hesperus' refer in sentence (2), the referent of the first will be identical with the referent of the second. Since that identity is known to hold *simply in virtue of knowing that sentence* (2) *is subject to the identity convention* the identity proposition expressed by (1) can convey no new information.

Here is a different way to make the same point. Suppose someone makes a statement to an audience *A* by writing down some sentence. In advance of *S* being written down, all that *A* knows is this. *S* states of some object that it stands in a two place relation to some, it may be the same, object. In addition *A* knows that the referring terms in *S* are subject to the identity convention.

Now let '□' represent a device, say a sticker, used to conceal a linguistic expression. When *S* is first written down the two place predicate in *S* is covered by '□'. So, when *S* is first written down, *A* initially sees:

S: '$(\exists x)$ 'Joan' refers to $x \rightarrow$(Joan □ Joan)'

What two place relation does *S* say that Joan stands in to Joan if 'Joan' has a referent? At this stage, *A* does not know. What *A* does know is that the identity convention holds for *S*. Hence, *A* knows that, whatever the two place predicate in *S* turns out to be when '□' is removed, *S* says that, if 'Joan' refers, the two place relation that predicate picks out holds between Joan and someone identical with Joan. Now consider the following two cases. In the first, ‘□’ is removed to reveal that *S* is:

S: ‘$(\exists x)$ ‘Joan’ refers to $x \rightarrow$(Joan likes Joan)’

S conveys something that *A* did not know prior to the removal of the obscuring device ‘□’. In the second case, before the removal of ‘□’ *A* sees the following sentence *S′* which is visually indistinguishable from the part of *S* *A* could see with ‘□’ still obscuring the two place predicate in that sentence:

S′: ‘$(\exists x)$ ‘Joan’ refers to $x \rightarrow$(Joan Joan)’

When ‘□’ is removed from *S′* *A* sees:

S′: ‘$(\exists x)$ ‘Joan’ refers to $x \rightarrow$(Joan = Joan)’

Prior to the removal of the device obscuring the two place predicate in *S′* the two tokens of ‘Joan’ on either side of that device are governed by the identity convention. Hence, prior to the removal of ‘□’ *A* knows that, if *S′* expresses the proposition its utterer intends it to express, the token of ‘Joan’ in *S′* preceding ‘□’ refers to something which is identical with the referent of its token succeeding that device.

Background knowledge of the identity convention governing *S′* gives *A* the following metalinguistic knowledge. No matter what the revealed two place predicate in *S′* turns out to be, if *S′* expresses the proposition its utterer intends it to express, the referent of the first token of ‘Joan’ in *S′* is identical with the referent of the other token of that name in the same sentence. Call that metalinguistic information *Im*.

The consequent of *S′* expresses an identity proposition. Knowledge of that proposition would give *A* information that is not metalinguistic. Call that information *I*. Suppose that *A* knows the antecedent of the proposition expressed by *S′*, $(\exists x)$ ‘Joan’ refers to x, is true. Will *A* acquire knowledge of its consequent, Joan=Joan, that is: knowledge of *I*, if *A* understands *S′* and knows that it expresses a truth? No. So, why not?

We have observed that having an item of knowledge can prevent one from acquiring an item of knowledge. One way this can happen is if the blocking knowledge is misleading. Another way is that one already has the knowledge that would otherwise be conveyed. The present case does not fall into either of these categories. The metalinguistic information *Im* is not misleading. In addition, knowledge of *I*, that is knowing the identity proposition which is the consequent of the proposition expressed by *S′*, does not amount to knowledge of *Im*. *Im* is metalinguistic. *I* is not. So, it remains to be explained why knowledge of *Im* prevents *A* from knowing *I* as a result of understanding *S′* and knowing it expresses a truth.

The explanation is this. Suppose, we are sitting in my office, and I utter the following sentence:

S: 'Joan is at location *L*'.

to inform about Joan's location. Now, suppose I impose the following stipulation. 'Location *L*' refers to a location just in case Joan is at that location. Call this the *location stipulation*, and suppose Joan is next door. Given the location stipulation *S* expresses a proposition that is true just in case Joan is next door. Nevertheless, even if Joan is next door, imposing the location stipulation renders an utterance of *S* uninformative. Knowledge that the location stipulation is in force disables *S* as a vehicle for conveying knowledge of the proposition it expresses.

Why does knowledge of the location stipulation have this effect? One who knows that 'next door' is subject to the location stipulation does not thereby know the proposition expressed by *S*. Nevertheless, what she does know, in virtue of knowing that 'location *L*' in *S* is governed by the location stipulation is this. If *S* expresses anything, it will express something true. But knowing the latter comes at a price. It prevents understanding *S* *alone* enabling one to know which proposition it expresses.

We can think of it like this. Consider a set of propositions $\{P_1, P_2,\ldots,P_n\}$ each member of the set specifies a location, different from that specified by any other member, that Joan could be at. Taken together the members of the set specify all the locations Joan could be at. Taken together the members of $\{P_1,P_2,\ldots,P_n\}$ exhaustively specify all Joan's possible locations.

What does someone know solely on the basis of understanding *S* and knowing it to be true? Only this. *S* expresses whichever member of $\{P_1,P_2,\ldots,P_n\}$ is true. What such a one does not know is which member of the relevant set of propositions is expressed by *S*. Moreover they lack that knowledge because understanding *S* implies knowing that the use of 'location *L*' is governed by the location stipulation.[19]

---

[19] What does it take to know which proposition is expressed by a sentence? That is a very good question that I will attempt to answer elsewhere. Without having an answer available here we can, at least, be confident of the following. If there are contingent identities, understanding an identity sentence of the form '$a=a$' will not enable one to know which proposition is expressed by that identity sentence. Here is the reason why this is so. Suppose that if Tib is identical with Tibbles, it is contingently so. In that case, as a matter of metaphysical possibility either of these situations could obtain:

In a like fashion knowing that the identity convention is in force disables 'Joan=Joan' from conveying knowledge of the identity proposition it expresses. One who knows the identity convention applies to 'Joan=Joan' thereby knows that, if it expresses anything, that sentence expresses a truth. The cost of so acquiring that knowledge is the same as in the previous case. What truth 'Joan=Joan' expresses requires a further input of information.

It needs to be emphasized that this application of the IC account to Frege's puzzle is limited to making room for the inflationist version of the no-difference solution to that puzzle. It is certainly not intended as a full blown defense of that solution. According to both the inflationist and deflationist versions of the no-difference solution sentence:

(1) Hesperus=Phosphorus

and sentence;

(2) Hesperus=Hesperus

are equally informative, and have the same epistemic status. The deflationist holds that the proposition expressed by sentence (1) is uninformative, and can be known *a priori*. The inflationist holds that the proposition expressed by sentence (2) is informative, and can be known only *a posteriori*. How can the inflationist explain away the apparent uninformativeness of sentence (2)? The IC account provides the explanation. We need to distinguish between what information would be gleaned by someone who understands sentence (2), and what information would be gleaned by coming to know the proposition expressed by sentence (2). Knowledge of the identity convention prevents one from knowing the proposition expressed by sentence (2) simply on the basis of understanding that sentence and knowing that it expresses something true.

Of course, this does not constitute anything like an adequate defense of the inflationist solution. For one thing we will be pushed towards the deflationist alternative of the no difference view if we combine taking belief to be a two place relation holding between a believer and a proposition with taking sentence (2) to express the same proposition as sentence (1). In any case, my main aim has not

---

Tib_1: Tib=Tibbles
$\neg(\text{Tib}_1)$: Tib is distinct from Tibbles.

If $\neg(\text{Tib}_1)$ obtains, 'Tibbles=Tibbles' will not express Tib_1. So, in order to know whether 'Tibbles=Tibbles' expresses Tib_1 one needs to know that $\neg(\text{Tib}_1)$ does not obtain. That is knowledge that can only be acquired *a posteriori* even by one who fully understands what linguistic conventions govern the use of 'Tibbles=Tibbles'.

been to present a complete solution to Frege's Puzzle. My main aim has been to show how the identity convention fosters an illusion about identity that lends unwarranted support to popular arguments against contingent and indeterminate identities.

That said, I will conclude by showing how the Identity Convention account helps us to see that Frege's Puzzle is, after all a puzzle about identity. Consider one of the examples given earlier that appears to show it is not. The same puzzle about differing informativeness arises for sentences that are not used to *state* identities such as sentences:

(3) Hesperus is the same size as Phosphorus.

and:

(4) Hesperus is the same size as Hesperus.

It should be clear what a defender of the IC account will say about sentences (3) and (4). Sentence (4) does not, to employ the terminology introduced earlier, satisfy the coincidence condition. Nevertheless, sentence (4), unlike sentence (3), does metalinguistically impute the identity relation which, evidently holds between *x* and *y* only if *x* is the same height as *y*. Hence, sentence (4) does metalinguistically impute a relation that, trivially, holds between a pair of relata only if they are the same size. As before, suppose that someone initially sees sentence (4) with a patch obscuring the relational predicate 'is the same size'. What that person sees is:

(4*) Hesperus ☐ Hesperus

Suppose before the patch is removed, the one viewing sentence (4*) knows that the two tokens of 'Hesperus' in that sentence are subject to the identity convention. Armed with this knowledge such a one may infer, before the patch is removed, that the referent of the left hand token of 'Hesperus' is identical with the referent of its right hand token in sentence (4*). Hence, such a one will be able to infer that, whatever the relational predicate in sentence (4*) turns out to be, the relation it attributes will, if sentence (4*) expresses a truth, hold between things of the same height. That is why, on the removal of the patch, learning what the contained predicate is does not impart new information about Hesperus.

If this is right, Frege's Puzzle turns out to be ineliminably a puzzle about identity in the following sense. A pair of sentences will pose the puzzle only if one member of the pair metalinguistically imputes the identity relation.

ANDRÉ GALLOIS

Syracuse University

THE SUBTRACTION ARGUMENT FOR METAPHYSICAL NIHILISM*

"Why is there something rather than nothing?" is a perennial question of metaphysics, if not the fundamental question as Heidegger maintains.[1] But what sort of thing does this question concern? Traditionally, the question is taken to concern concrete objects: why is there something concrete rather than nothing concrete.[2] If there had to have been something concrete, as some argue, for example, E.J. Lowe,[3] then the question receives a trivial answer. And perhaps the question does admit of a trivial answer, but surely not this trivial answer for it seems intuitively compelling that there might have been nothing concrete. But is there anything more than bare intuition to support the view that there might have been nothing concrete, a view that Lowe has termed *metaphysical nihilism*? That is, can the metaphysical nihilist respond to Lowe's argument with anything more than a Lewisian blank stare? It seems she can, for Thomas Baldwin has proposed his own argument,[4] subsequently refined by Gonzalo Rodriguez-Pereyra,[5] on behalf of the metaphysical nihilist. Baldwin terms this argument the *subtraction argument*, an intuitively compelling description which one must take seriously, but not too seriously. This argument, we maintain, is a cogent argument which should constrain theories of modality. Thus, we reject David Armstrong's claim that whether there might have been nothing is a question best decided *after* deciding on the best theory of modality: determining whether or not the empty world is possible is not "spoils for the victor" as

---

* The authors would like to acknowledge support from the Arts and Humanities Research Council for a Research Leave Award and from the University of York for an Anniversary Lectureship. We would like to thank audiences at a Modality Workshop at the Arché Research Centre in St. Andrews, the University of Keele and the University of York where portions of this paper were presented. The authors are grateful to Alexander Paseau, Gonzalo Rodriguez-Pereyra, and Timothy Williamson for helpful discussions and correspondence on the subtraction argument.

[1] *An Introduction to Metaphysics* (New Haven: Yale, 1959), pp. 7–8.

[2] Just what is meant by 'concrete' we explain below.

[3] "Why Is There Anything at All?" *Aristotelian Society*, Suppl. Vol. LXX (1996): 111–20, *The Possibility of Metaphysics: Substance Identity and Time* (New York: Oxford, 1998), pp. 252–55, "Metaphysical Nihilism and the Subtraction Argument," *Analysis*, LXII (January 2002): 62–73.

[4] "There Might Be Nothing," *Analysis*, LVI (October 1996): 231–38.

[5] "There Might Be Nothing: The Subtraction Argument Improved," *Analysis*, LVII (July 1997): 159–66.

0022-362X/05/0206/303–25

Armstrong puts it, but rather helps to determine which theory is the victor.[6]

In establishing the cogency of the subtraction argument, we first show that Baldwin's argument contains an ambiguous premise, one that gives rise to four interpretations. Two of the interpretations, which correspond to the readings given by Alexander Paseau,[7] render the argument invalid, but they are seemingly wilful misreadings of Baldwin's premises since they are inconsistent with the sort of reasoning driving the subtraction argument. A third interpretation, which corresponds to the reading given by Rodriguez-Pereyra,[8] renders the subtraction argument valid, but question begging because it renders one of the premises as tantamount to metaphysical nihilism itself. Finally, a fourth interpretation of the subtraction argument, which we offer, renders the argument valid, and, in our view, persuasive. It is persuasive because it brings out the fact that the subtraction argument is an argument not primarily about possible worlds, in and of themselves, in which case the subtraction argument would indeed be question begging, but rather it is an argument that concerns primarily concrete individuals and their modal properties. In particular, it concerns whether one such modal property, namely, subtractability, which we characterize precisely, is an intrinsic property of concrete individuals. On this interpretation there is then no question of the argument being question begging. In addition, it provides a way of settling whether there might have been nothing concrete not by appeal to a pre-existing theory of possible worlds, as Armstrong[9] and David Lewis[10] do, but rather by appeal to whether concrete objects have a certain modal property.

WHAT EXACTLY IS THE SUBTRACTION ARGUMENT?

Baldwin's argument contains the following three premises.

(A1) There might be a world with a finite domain of 'concrete' objects.
(A2) These objects are, each of them, things which might not exist.
(A3) The nonexistence of any one of these things does not necessitate the existence of any other such things (*op. cit.*, p. 232).

---

[6] *A Combinatorial Theory of Possibility* (New York: Cambridge, 1989), p. 64.

[7] "Why the Subtraction Argument Does Not Add Up," *Analysis*, LXII (January 2002): 73–75.

[8] "Metaphysical Nihilism Defended: Reply to Lowe and Paseau," *Analysis*, LXII (April 2002): 172–80.

[9] *A Combinatorial Theory of Possibility: A World of States of Affairs* (New York: Cambridge, 1997).

[10] *On the Plurality of Worlds* (New York: Blackwell, 1986).

From these premises, Baldwin argues for metaphysical nihilism in the following way.

> By (A1), starting from the actual world W, there is an accessible world w_1 whose domain of concrete objects is finite. Pick any member x_1 of this domain: by (A2) there is a world accessible from w_1, w_2, which is just like w_1 except that it lacks x_1 and any other things whose non-existence is implied by the non-existence of x_1. Since, by (A3), the domain of w_2 does not contain things which do not exist in w_1, it follows that the domain of w_2 is smaller than that of w_1. This procedure of subtraction can then be iterated, until we get to a world w_{min} whose domain consists of one or more concrete objects, such that the non-existence of one implies the non-existence of all. By (A2) the non-existence of one of these objects is possible, so there is a world w_{nil} just like w_{min} whose domain lacks all these objects; and since, by (A3), the non-existence of these things does not require the existence of anything else, w_{nil} is a world in which there is no concrete object at all. If one now allows that accessibility between worlds is transitive (the characteristic S4 assumption) it follows that w_{nil} is accessible from, or possible relative to, the actual world (*op. cit.*, p. 232).

Baldwin's phrase, 'this procedure of subtraction', is undoubtedly meant to be a metaphor. If it is taken literally, it suggests a physical operation on a possible world to create another possible world by taking away some of the objects in the original possible world. Because possible worlds are not the kind of thing one can create by a physical operation, such a literal interpretation of 'this procedure of subtraction' is thus inadvisable and was surely never intended. Instead of taking subtraction in such a literalist sense, we should try to understand subtraction in terms of a relation between the objects which exist at different worlds. In this way, we can make sense of subtracted worlds, that is, worlds whose domains stand in a certain relation to each other, a relation suggested by Baldwin's subtraction metaphor. In evaluating the validity of Baldwin's argument, it will prove important to pay attention to this metaphor.

In a recent note (*op. cit.*), Paseau has argued that (A1)–(A3) cannot support Baldwin's reasoning because, in his view, (A3) admits of two readings, neither of which, together with (A1) and (A2) entail metaphysical nihilism.[11] The first reading of (A3) that Paseau distinguishes is that "the non-existence of any given one of the finitely many objects of which (A1) and (A2) speak, namely, $x_1, x_2, \ldots, x_n$, does not necessitate the existence of any other given one of these x_i" (*op. cit.*, p. 74). On this reading, the following set of worlds expressed as sets of concrete objects is a counterexample to the validity of the subtraction argument: $\{x_1\}$, $\{x_2\}$, $\{o\}$, $\{x_1, x_2\}$, $\{x_1, o\}$, $\{x_2, o\}$, $\{x_1, x_2, o\}$. Paseau's second reading of (A3) is that "the non-existence of any one of these x_i does not necessitate that there is even one of the x_i" (*op. cit.*, p. 74). And the counterexample for his first reading of (A3) also works for this reading. Thus, Paseau concludes, the subtraction argument is invalid.

---

[11] Paseau's explicit target is Rodriguez-Pereyra's summary of Baldwin's reasoning. See Rodriguez-Pereyra, "There Might Be Nothing: The Subtraction Argument Improved," p. 164.

Paseau's thought seems to be this. There is a world with a finite number of contingently existing concrete objects, $\{x_1, x_2\}$. On both readings, (A3) requires that there is a world at which neither x_1 nor x_2 exists. But in his model, it is true that for every ordered pair consisting of one of the x_i and another concrete object o, there is a world at which neither exists. So (A3) is true for every one of the x_i on both readings, and this holds whether or not o is one of the x_i, that is, whether o is an element of the world specified in (A1), but there is no world containing no concrete objects at all. Consequently, though (A1), (A2), and Paseau's readings of (A3) are true in his model, metaphysical nihilism is not.

While we grant that Paseau has distinguished two readings of (A3) on which the "subtraction" argument is invalid, neither of these readings can be what Baldwin originally intended, because neither respects the subtraction metaphor. The point of (A3) is to capture the intuition that a concrete, contingent object can always be *subtracted* from a world in which it exists without necessitating the existence of any distinct concrete object whatsoever at the resulting world. If we try to use the subtraction metaphor to understand Paseau's interpretation of the argument, that is, to ask what subtraction relation holds between the worlds of Paseau's model, it seems that some world, w_2, is the "result" of subtracting some object x_1 from w_1 iff everything other than x_1 which exists in w_1 also exists in w_2. Given this understanding of subtraction, the world $\{x_2, o\}$ is a result of subtracting x_1 from $\{x_1, x_2\}$, and the world $\{o\}$ is a result of subtracting x_1 from $\{x_1\}$. This suggests the following definition of the "subtraction" relation between worlds:

$$\text{(PS)}\ \forall w_1 w_2\,(Sw_1w_2 \equiv \exists x(E!xw_1 \wedge \neg E!xw_2 \wedge \forall y((x \neq y \wedge E!yw_1) \supset E!yw_2)))^{12}$$

---

[12] Variables 'x' and 'y' range unrestrictedly over concrete objects and variables 'w_1' and 'w_2' range over worlds, and '$E!$' is the exists-at-a-world predicate. Some model theorists may be concerned that there is no set large enough to be the domain of the concrete object quantifiers in this formula. However, after Richard Cartwright's attack on the 'All-in-one Principle', the claim that the domain of quantification is a set should be seen controversial. In our view, the intelligibility of the quantification in (PS) and the other formulae in this paper is more obvious than the All-in-one

Reading subtraction in Paseau's way makes the operation of subtracting from one collection of things consistent with adding from a distinct collection of things. For example, on this reading of subtracting, John can go from having five oranges and no apples to having three oranges and one apple by the operation of subtraction alone. In contrast, the intuitive notion of subtraction is inconsistent with such covert addition. It is clear that Paseau has not captured the intuition behind Baldwin's original argument because he has paid *too little* attention to the metaphor of subtraction.

There is another reading of (A3) that Paseau does not consider. This reading is that the nonexistence of any of the concrete objects of which (A1) and (A2) speak does not necessitate the existence of some concrete object or other. In Paseau's model the nonexistence of x_1 necessitates a disjunction of existential claims: either x_2 exists or o exists. But this disjunction entails that something exists, so the nonexistence of x_1 necessitates that something exists. If we read the second 'any' in (A3) as having narrow scope, his model is no longer a counter model. Thus (changing 'any' to 'some' to clarify the scope):

> (A3*) The nonexistence of any one of these things does not necessitate that some other concrete object, from the domain of any world whatsoever, exists.

Taking w_1 to be the world mentioned in (A1), we can formalize this as:

(Q-B) $\forall x(E!xw_1 \supset \neg\forall w_2(\neg E!xw_2 \supset \exists y(y \neq x \wedge E!yw_2)))$[13]

The drawback of reading (A3) in this way is that the subtraction argument is now question begging. In order to show this, we formalize metaphysical nihilism as the claim:

(MN) $\exists w\forall x\neg E!xw$

Now (Q-B) is logically equivalent to the claim:

(Q-B*) $\forall x(E!xw_1 \supset \exists w_2(\neg E!xw_2 \wedge \forall y(y \neq x \supset \neg E!yw_2)))$

(Q-B*) would be vacuously true if nothing existed at w_1. Otherwise, it is only true if, for each thing that exists at w_1, there is a world at which neither it nor anything else exists. So (Q-B) is equivalent to the disjunction: either w_1 is the empty world, or for each thing that

---

Principle. For Cartwright's attack on the All-in-one Principle see his "Speaking of Everything," *Noûs*, XXVIII (March 1994): 1–20.

[13] Paseau's reading of (A3) can be formalized as:

(P) $\forall x(E!xw_1 \supset \neg\exists y\forall w_2(\neg E!xw_2 \supset E!yw_2))$

exists at w_1, there is an empty world. So, if (A3) is read as (Q-B) we get a question-begging argument for metaphysical nihilism. This reading, then, is obviously not desirable. It seems, though, that this reading is intended by Rodriguez-Pereyra in his revision of (A3). Writing in response to Paseau, Rodriguez-Pereyra reads (A3) in the following way:

> (A3**) The non-existence of any of the x_i that exist in w_1 does not necessitate the existence of any other concrete* object, whether or not these exist in w_1. That is: for all worlds w and for all the concreta* x_i in w_1, if x_i exists in w then if there is a world w^* where x_i does not exist, then there is a world w^{**} where the only existing concreta* are those of w except x_i (i.e. w^{**} is such that for every concrete* object y, y exists in w^{**} if and only if $y \neq x_i$ and y exists in w).[14]

Rodriguez-Pereyra's version of (A3) consists in a statement and a gloss of it.[15] These are not equivalent. The statement is equivalent to (A3*) and so (Q-B) appears to be an adequate symbolization of it. But Rodriguez-Pereyra's gloss on it, the claim that follows "That is:" is not symbolizable as (Q-B) or any other claim logically equivalent to (Q-B). The gloss seems to call for the following symbolization:

$$\text{(R-P)}\ \forall x(E!xw_1 \supset \forall w_2(E!xw_2 \supset (\exists w_3 \neg E!xw_3 \supset \exists w_4 \forall y(E!yw_4 \equiv (y \neq x \wedge E!yw_2)))))$$

(R-P) is not equivalent to (Q-B) as it, on its own, does not entail the existence of the empty world. And the argument from (A1), (A2), and (R-P) we are willing to concede is valid. So we have a candidate reading of (A3) that makes the subtraction argument potentially persuasive: it is valid and it is not question begging. But (R-P) has its drawbacks. To begin, it is overly complicated. The intuitive idea behind (A3) is that subtraction is an operation that removes things from a collection but does not add anything more, and that we can always subtract contingent concrete objects from worlds. This condition is met if the resulting collection lacks the subtracted object and contains nothing which is not in the original collection. Call this a *subtraction world* of the original. What (A3) really needs to assert is that for every object at w_1, there is a subtraction world which lacks that object, and for each object at each of those subtraction worlds, there is a subtraction world which lacks that object, and so on. Since (A3) does not depend upon there being only finitely many objects in the original world, this idea can be formalized much more simply than (R-P) does. A straightforward symbolization of this idea is the following:

---

[14] "Metaphysical Nihilism Defended: Reply to Lowe and Paseau," p. 172.

[15] We explain what Rodriguez-Pereyra means by 'concrete* object' in footnote 27.

(B) $\forall w_1 \forall x(E!xw_1 \supset \exists w_2(\neg E!xw_2 \wedge \forall y(E!yw_2 \supset E!yw_1)))$

(B) captures Baldwin's original intention when he wrote: "Since, by (A3), the domain of w_2 does not contain things which do not exist in w_1" (*op. cit.*, p 232). It is clear that (B) is not equivalent to (R-P) because (B) uses the ancestral of the subtraction relation and (R-P) does not.[16] The effect of using the ancestral of the subtraction relation is that (A1), (A2), (B) entails metaphysical nihilism without requiring the existence of a singleton world, where the use of (R-P) in place of (B) does require the existence of the singleton world.[17] Consequently, (B) allows for the existence of symbiotic objects, that is, pairs of objects such that if one goes out of existence so does the other, whereas (R-P) does not. In our view, it seems that such objects should not be ruled out, and Baldwin evidently intended that such objects not be ruled out when he wrote: "by (A2) there is a world accessible from w_1, w_2, which is just like w_1 except that it lacks x_1 and *any other things whose non-existence is implied by the non-existence of* x_1" (*op. cit.*, p. 232, emphasis added). In fact, mereological fusions of concrete objects provide clear examples of objects which cannot be subtracted from a world without subtracting something else, and if at least some mereological fusions of concrete objects are themselves concrete, which seems overwhelmingly plausible, then (R-P) is false. It seems then, that (A3) should be understood as (B), since (B) renders the subtraction argument valid but not question begging.[18] And though it might be challenged, (B) is backed up by a basic modal intuition: *if there are some concrete objects, there could have been fewer of them.* This intuitive claim is *a priori,* and so it is necessarily true if true at all.[19] Hence, we have a valid argument from premises supported by intuition,[20] and that gives us good reason to believe the conclusion.

---

[16] Not strictly the ancestral, but a relation which is co-extensive with the ancestral relation when the domain of w_1 is finite, as it is. Thanks to an anonymous referee for pointing this out.

[17] In fact, (A1) and (B) entail (MN) without (A2). We prove this in the appendix.

[18] It may be suggested, though, that (B) is not a reading of (A3) at all, but, rather, a replacement of it. This may be so. What we have attempted to do is give an adequate formalization of the reasoning Baldwin uses to argue for metaphysical nihilism. It may be that he himself did not capture this reasoning with (A3).

[19] There are of course exceptions to the rule that *a priori* claims are necessary, but the present claim does not seem to be one of those exceptions.

[20] Of course, the claim that, necessarily, if there are some concrete objects, there could have been fewer of those concrete objects, is not equivalent to (B), as Paseau's counterexample makes clear. However, a case can be made that the sentence intuitively accepted as true is a natural expression of (B), for when specifying a possibility,

TO WHAT DOES THE SUBTRACTION ARGUMENT APPLY?

In the next section we will be arguing that ordinary unphilosophical intuitions about concrete objects do support the premises of the subtraction argument. But first we need to specify exactly the class of objects to which the subtraction argument is intended to apply, a class we have up to now been picking out with the term 'concrete'. 'Concrete' is a term of art variously used to specify different, but overlapping sets of objects in different philosophical contexts, for example, spatiotemporal objects, causally efficacious objects, middle-sized dry goods. Without specifying which set of objects 'concrete' picks out, it is dubious that we have any ordinary unphilosophical intuitions about concrete objects as such, because we are not sure just what sort of object we are considering. Since it is our aim to show that there is a sound version of the subtraction argument which does not rely upon any substantive metaphysical theses, it is important that the premises are founded on clear cases of ordinary unphilosophical intuitions, so we need to specify which particular set of objects 'concrete' picks out in this context.

To do this, we need to consider the sort of object of which we would, in our most unphilosophical moments, say there might have been none, or equivalently, for present purposes, the sort of object one has in mind when one considers why there is something rather than nothing. And what sort is this? Well, it is the sort that includes cabbages, kings, pieces of sealing wax, and the like, and excludes numbers, propositions, properties, and the like (if there are any, of course). What is it that the former group has in common and the latter group lacks? It seems to us that all of the former, and none of the latter, are such that:

(a) they exist at locations in space-time, and
(b) they have some intrinsic[21] qualities.

---

it is a conversational implicature that all pertinent deviations from actuality are mentioned. Thus, someone who was thinking of a possibility in which there were fewer of the actual concrete objects, but not fewer concrete objects altogether, would be violating H.P. Grice's maxim of informativeness if he described it simply as a world in which there are fewer of the actual concrete objects. We claim that (B) is doing no more than making precise the thought which someone would be intending to express in a normal conversational context should she assert that, necessarily, if there are some concrete objects, there could have been fewer of them. See Grice, "Logic and Conversation," in Peter Cole and Jerry L. Morgan, eds., *Syntax and Semantics 3: Speech Acts* (New York: Academic Press, 1975), pp. 41–58.

[21] Defining 'intrinsic' is no easy matter—witness the debate over whether 'intrinsic' can be defined in terms of modal independence. See, for instance, Jaegwon Kim, "Psychophysical Supervenience," *Philosophical Studies*, XLI (January 1982): 51–70; Lewis, "Extrinsic Properties," *Philosophical Studies*, XLIV (September 1983): 197–200; Rae Langton and Lewis, "Defining 'Intrinsic'," *Philosophy and Phenomenological Re-*

and these two characteristics are together the most salient for the subtraction argument. So, for present purposes, we fix the interpretation of 'concrete' as applying to all and only those objects which satisfy both (a) and (b). Hence, metaphysical nihilism, most precisely stated, is the view that there might have been nothing which has both spatiotemporal location and some intrinsic qualities.

To test whether we have identified the sort of object that concerns us when we, in our unguarded moments, say that there might have been nothing of that sort, we should consider some cases that might seem to tell against our account. For example, the account includes as concrete some things which are not obviously so, such as rainbows. But though it might seem odd to classify rainbows as concrete, it does give the right result for the subtraction argument. In this context, we are using 'concrete' to pick out the sort of objects of which there might have been none, and rainbows surely fall into this category: if a possible world consisted only of rainbows, it seems that world would contain something rather than nothing.[22] So, rather than showing the account to be wrong, rainbows show the account to be right. Similarly, the account classifies electric fields as concrete and that is correct, for a world containing nothing but electric fields would contain something rather than nothing.[23]

---

search, LVIII (June 1998): 333–45; Dan Marshall and Josh Parsons, "Langton and Lewis on 'Intrinsic'," *Philosophy and Phenomenological Research*, LXIII (September 2001): 346–51; Langton and Lewis, "Marshall and Parsons on 'Intrinsic'," *Philosophy and Phenomenological Research*, LXIII (September 2001): 353–55; Ted Sider, "Maximality and Intrinsic Properties," *Philosophy and Phenomenological Research*, LXIII (September 2001): 357–64; Brian Weatherson, "Intrinsic Properties and Combinatorial Principles," *Philosophy and Phenomenological Research*, LXIII (September 2001): 365–80; Lewis, "Redefining 'Intrinsic'," *Philosophy and Phenomenological Research*, LXIII (September 2001): 381–98; John Hawthorne, "Intrinsic Properties and Natural Relations," *Philosophy and Phenomenological Research*, LXIII (September 2001): 399–403. Indeed, the attempt to define 'intrinsic' in terms of modal independence has all the hallmarks of a degenerating research program. But a salient feature of this debate is the general agreement over *which* properties are intrinsic. This should give us confidence that the concept of intrinsicality is in good order and that it can be put to philosophically useful purposes, even if we struggle to explicate it.

[22] Of course, a possible world consisting of only rainbows is *physically* impossible, but we are concerned here with the broadest, metaphysical notion of possibility. Similarly, if we allow that ghosts are metaphysically possible, they would be concrete for they have both locations and intrinsic properties, despite the fact that ghosts, it seems, do not exclude anything. But this again seems right: it chimes with the thought that a world consisting of nothing but ghosts would be a world in which there was something rather than nothing.

[23] Indeed, Michael Faraday proposed that the universe contains only fields and no matter. In proposing this, he was surely not proposing that the universe consists of nothing of the sort relevant to the question, 'Why is there something rather than nothing?' See his "A Speculation Touching Electric Conduction and the Nature of Matter," in *Experimental Researches in Electricity*, Volume 2 (London: Richard and John

Another case that, on first reflection, might seem to pose a difficulty for the account is spacetime points. Spacetime points are spatiotemporally located, and so satisfy (a), but they fail to satisfy (b), that is, they have no intrinsic properties.[24] So they are not concrete but are abstract according to our definition. And this too seems right, since a world which contained only spacetime points would contain nothing which had any intrinsic properties, and a world like that would be a world which might as well contain nothing at all from the perspective of the question, 'Why is there something rather than nothing?'. So rather than showing the account to be wrong, spacetime points show the account to be right.

The definition of concreteness in terms of (a) and (b) differs from that to be found in Baldwin (*op. cit.*, p. 233) and Rodriguez-Pereyra,[25] who both think that the key mark of concreteness is failure to satisfy nonvacuously the identity of indiscernibles. If we assume that no intrinsic properties are such that their instantiation at one point in space-time places a restriction upon where else in space-time they might be instantiated, it is plausible that all objects existing in space-time and possessing intrinsic qualities do fail the identity of indiscernibles. But the reverse inference turns on the thought that the only way to understand the possibility of distinct but indiscernible duplicates is to think of those objects as existing at different locations in space-time. This may be true, but to argue the point would take us way beyond the scope of the subtraction argument. We conclude that the conjunction of (a) and (b) is either equivalent to Baldwin and Rodriguez-Pereyra's definition of concreteness in terms of nonvacuous failure to satisfy the identity of indiscernibles, or preferable to it because it relies on fewer metaphysical assumptions.

However, the current definition does face one problem raised by Rodriguez-Pereyra: extended objects each have an infinite number of overlapping spatial parts and if each of those is concrete, then there can only be a finite number of concrete objects if there are no *extended* concrete objects.[26] By the definition we are proposing, an

---

Edward Taylor, 1844), pp. 284–93.

[24] Such clearly abstract objects as centres of gravity, though they have location, also lack intrinsic properties.

[25] "There Might Be Nothing: The Subtraction Argument Improved," p. 165.

[26] See "There Might Be Nothing: The Subtraction Argument Improved," p. 163. Rodriguez-Pereyra credits Cartwright as the inspiration for the argument, who in turn credits Leibniz. See Cartwright, "Scattered Objects," in his *Philosophical Essays* (Cambridge: MIT, 1987), pp. 171–86, and Leibniz, "Fourth Letter to Clarke," in Leroy Loemker, trans. and ed., *Leibniz: Philosophical Papers and Letters*, 2nd ed. (Boston: Reidel, 1970).

infinite number of these spatial parts of a concrete object will be concrete, for they will exist in spacetime and have intrinsic qualities, so the first premise of the subtraction argument would be false. Of course, it is not necessary that all an object's spatial parts have intrinsic qualities, but that does not stop an infinity of them being concrete.[27]

To make the problem graphic, consider a possible world which contains a 1 meter square white card with a 10 centimeter red circle in the middle. This is spatially located and has at least one intrinsic property (the red area). There are an infinite number of circles of card with a diameter between 10 centimeters and 50 centimeters, each of which has a spatial location and an intrinsic property. So if one such card exists, an infinite number of concrete objects exists.

Now, it is quite clear that the unphilosophical intuition that there could have been a finite number of concrete objects would not be swayed by the alleged counterexample of the infinite number of overlapping spatial parts of a finitely extended object. The problem we face is to articulate this concept of a concrete object in a way which shows why Rodriguez-Pereyra's objection is irrelevant, and to do so while maintaining the intuition that if there are some, there could have been fewer. Thus, requiring that concrete objects exclude others from their locations would solve the problem of too many spatial parts but rule out rainbows and electric fields, which we noted above that we wanted to include.

It might seem that one solution is to appeal to the sortal-dependence of identity and thus of number. For example, one might say that the infinitely many concentric circular regions contained within the 1 meter square sheet of card are none of them *pieces of card.* Of course, if the square card was cut into a circle, there would then be a piece of card coincident with one of them, but only with one of them. Furthermore, there is no other sortal term which applies equally to all those regions. So when we count by sortals, we count one square piece of card and ignore all the circular sub-regions. But this solution will not do, and not only because of the enormous difficulty in distinguishing sortal terms from other common nouns. For if we took this tack, then metaphysical nihilism would be the thesis that it is possible that, for every sortal *S*, there is nothing which is an *S*, and universal quantification over sortals is not easy to understand.

---

[27] Rodriguez-Pereyra's version of this argument relies on the claim that the parts of a concrete object are just as concrete as the object itself. If there are mereological fusions of concrete and abstract objects, this is not obviously true and is certainly not true on his definition of a concrete object, for a fusion will fail to satisfy the identity of indiscernibles if just one of its parts does. See "There Might Be Nothing: The Subtraction Argument Improved," p. 163.

Rodriguez-Pereyra's solution to the problem is to restrict the subtraction argument to concrete* objects, which are not merely concrete but also maximal, memberless occupants of a connected region. Having made this move, it is no longer possible to appeal to unphilosophical intuition to support the premises of the argument.

Consider again the possible world which contains just the white card with a red circle on it. We would say that in such a world there is a square and there is a circle, the red one. But we would be inclined to deny that there is, in such a world, also a white circle with a red circle on it. Here is Ted Sider making the point with a three-dimensional example:

> Consider a solid cube of gold with one-meter edges. This cube has a spherical part, S, one meter in diameter. Although S is *spherical*, linguistic intuition assures us that it is no *sphere*. To be a sphere, it would have to be extracted from the cube in which it is embedded. Thus, *being a sphere* is border-sensitive (*op. cit.*, p. 358).

The intuition here has nothing to do with whether *being a sphere* is a sortal concept or not. Rather the point seems to be that where there is a sphere, there must be a (spherical) *boundary* between that thing and what is adjacent to it. Suppose that Sider's cube was not solid gold, but contained a spherical lump of platinum. Then linguistic intuition would allow the platinum to be a sphere, despite its being embedded in a gold cube. The difference is that there is a natural boundary between the sphere of platinum and the gold which encases it, but there is no such boundary between the "sphere" of gold and the gold which encases it.[28] And this is precisely what seems wrong with the infinity of white circles contained within the white card: none of them have a natural boundary and as a result, none are relevant to the unphilosophical judgement about how many concrete objects there are. So we propose the following severally necessary and jointly sufficient conditions on being a concrete object:

(a) it exists at a location in spacetime,
(b) it has some intrinsic quality, and
(c) it has a natural boundary.

---

[28] This is not quite what Sider means by 'border-sensitive'. On his view borders are easy to have and the spherical lump does have a border. The point is that whether the spherical lump of gold is a sphere depends in part upon what is outside its border. We are generalizing this thought: boundaries are borders, but whether a border is a boundary or not is itself border-sensitive. It is possible that the border-sensitivity of 'boundary' diverges from the border-sensitivity of 'sphere', so that some spheres lack boundaries and some bounded spherical objects are not spheres. If there are spheres which lack boundaries, then not all spheres are concrete, even if they have location and intrinsic qualities. Hence, someone who denied Sider's linguistic intuition about 'sphere' could still accept our basic point about boundaries.

We take the concept of a natural boundary to be primitive, at least in the context of the subtraction argument.[29] As Sider's example shows, one sort of boundary occurs when the stuff within the boundary is of a different natural kind from the stuff outside the boundary. But it may be that a difference in natural kind is not necessary if there is some other significant difference in kind. Thus there appears to be a natural boundary between the legs and top of a table when these are made from different pieces of wood but not when the whole table is carved from a single tree-trunk (or moulded in plastic). Whether there is a natural boundary between the red circle and the rest of the white card in our example is less clear, but just so long as there are only a finite number of natural boundaries within any finitely extended concrete object, this proposal avoids the Rodriguez-Pereyra objection.

Clearly, mereological fusions of contiguous objects can have a natural boundary and thus some mereological fusions of concrete objects are themselves concrete. Whether mereological fusions of noncontiguous concrete objects count as concrete is a matter we intend the definition to leave open by taking 'natural boundary' as primitive, but there does not seem to be any intuitive cost in giving a negative answer.

Potentially more problematic is the fact that few, if any, ordinary concrete objects have precisely defined boundaries. It is this which leads to the so-called Problem of the Many: a cloud consists of a set of water droplets, but for each such set there is another set which has one more or one fewer droplet and just as good a claim to constitute a cloud. So now it seems that where we thought there was just one cloud in the sky, there are millions.[30] This would only present a problem for the subtraction argument if imprecise boundaries created an infinitude of concrete objects. However, in the cloud example this does not seem to be the case, because there is a finite number of water droplets and thus a finite number of ways of making the boundary of the cloud precise.

It seems, then, that given our definition of concreteness and the undeniable fact that many concrete objects have imprecise boundaries, the first premise of the subtraction argument, that there might be a finite number of concrete objects, can only be true if at some level, however microscopic, the world might have consisted of items with precise natural boundaries. Whether it does is, of course, an empirical matter. What (A1) commits us to, on the current account, is the possibility that there should be precise boundaries in nature. And as a mere possibility, that is unexceptionable.

---

[29] This is not an entirely uncontroversial assumption as some deny that there are any boundaries, and others think they inherently involve some arbitrariness. But the issues which drive this debate over boundaries seem to be peripheral to the debate over whether there might have been nothing, and so it seems we can safely put them to one side and take natural boundaries as primitive. For developments on the debate over boundaries, see Barry Smith and Achille Varzi, "Fiat Boundaries and Bona Fide Boundaries," *Philosophy and Phenomenological Research*, LX (March 2000): 401–20.

[30] For the original formulation of this problem, see Peter Unger, "The Problem of the Many," *Midwest Studies in Philosophy*, V (1980): 411–67.

SHOULD WE ACCEPT THE SUBTRACTION ARGUMENT?

We have given a formally valid argument for metaphysical nihilism, namely:

$$\begin{array}{ll} \text{(A1)} & \exists w \exists x \exists y ((E!xw \wedge E!yw) \wedge \forall z (E!zw \supset (z = x \vee z = y))) \\ \text{(B)} & \forall w_1 \forall x (E!xw_1 \supset \exists w_2 (\neg E!xw_2 \wedge \forall y (E!yw_2 \supset E!yw_1))) \\ \hline \text{(MN)} & \exists w \forall x \neg E!xw \end{array}$$[31]

We then explained what the quantifiers in this argument range over, namely objects which exist at locations in space-time, have some intrinsic qualities and have natural boundaries. Given this interpretation of the subtraction argument, are its premises true? Each of the premises is supported by a modal intuition, namely:

(1) There could be a finite number of concrete objects.
(2) Necessarily, if there are some concrete objects, there could have been fewer of those concrete objects (and no other concrete objects).

(1) is a statement of what looks to be a secure modal intuition. To deny (1) is to assert that it is impossible that there is a finite number of concrete objects. So if the plausibility of metaphysical nihilism rests on (1), then it rests on firm intuitive ground for surely it is more plausible to suppose that there might have been no concrete objects than to suppose that it is *impossible* that there is a finite number of such objects. It thus seems unpromising to accept that the argument is valid but deny (1). Consequently, in order to maintain that metaphysical nihilism is false, it seems that one must either (i) deny (2), or (ii) accept (2) but claim that (2) does not support a claim as strong as (B), which is needed for the subtraction argument. We turn first to those who deny (2).

Claim (2) would find opposition from Bernard Linsky and Edward

---

[31] Without loss of generality, we assume that there might have been two concrete objects. See the appendix for details.

Zalta[32] and Timothy Williamson[33] who maintain that everything, including everything concrete, is a necessary existent.[34] On their view, which we dub *LZW*, there are some concrete objects, the concrete objects that exist in the actual world, but it is not possible that there are only fewer of those concrete objects since each of those objects is a necessary existent. But also on this view, everything concrete is merely contingently so.[35] This component of the view provides a way of reconciling it with metaphysical nihilism and a version of claim (2) consistent with its motivation. Since concrete objects are only contingently concrete, there is no opposition from LZW to the claim that there is a world containing no objects that are concrete in that world, which is just what metaphysical nihilism maintains, as long as every object that exists in this world, including the concrete objects, exist in that world. Thus, LZW and metaphysical nihilism are consistent as stated. Modification, though, is necessary to render claim (2) consistent with LZW, but the modification required is entirely consistent with the motivation behind (2). The following version of (2) is consistent with both LZW and the intuition behind (2).

> (2*) Necessarily, if some objects are concrete, then it could have been that fewer of those objects are concrete (and no other objects are concrete).

(2*) together with (1) entails that there might have been no objects that are concrete, which is what metaphysical nihilism maintains. Thus, Linsky, Zalta, and Williamson can be metaphysical nihilists of a sort, even though they maintain that everything, including everything concrete, is a necessary existent.

Lowe presents a different case against (2), one that does not allow him to be a metaphysical nihilist of any sort.[36] He argues that since some abstract objects exist necessarily and abstract objects depend for their existence upon there being concrete objects, it is necessary that there are concrete objects.[37] Thus, according to Lowe, metaphysical nihilism is false. Lowe grants that Baldwin's argument is valid, and he grants that (A1), and so, presumably (1), is true. Lowe denies (A3), and so presumably (2). Lowe's argument depends crucially on four premises, which are open to serious doubt.

---

[32] "In Defense of the Simplest Quantified Modal Logic," in J.E. Tomberlin, ed., *Philosophical Perspectives 8: Issues in Logic and Language* (Atascadero CA: Ridgeview, 1994), pp. 431–58, "In Defense of the Contingently Nonconcrete," *Philosophical Studies*, LXXXIV (December 1996): 283–94.

[33] "Necessary Identity and Necessary Existence," in Rudolf Haller and Johannes Brandl, eds., *Wittgenstein: Towards a Re-Evaluaction: Proceedings of the 14th International Wittgenstein-Symposium*, Volume I (Vienna: Holder-Pichler-Tempsky, 1990), pp. 168–75, "Bare Possibilia," *Erkenntnis*, XLVIII (1998): 257–73, "Necessary Existents," in Anthony O'Hear, ed., *Logic, Thought, and Language* (New York: Cambridge, 2002), pp. 233–52.

[34] Linsky and Zalta and Williamson would also accept that it is necessary that everything is a necessary existent which has the effect that if we choose an arbitrary possible world containing concrete objects, everything in that world is a necessary existent, and so, relative to that world, it is not possible that there are only fewer of those concrete objects.

[35] On this theory, abstract objects are necessarily nonconcrete objects.

(L1) The only abstract objects are sets and universals.
(L2) Abstract objects exist necessarily.
(L3) The empty set is impossible.
(L4) It is possible that *F*s are generically existentially dependent on *G*s while *G*s are identity dependent on *F*s just in case the *F*s have ontological priority over the *G*s.

(L1) and (L2) are only dubiously consistent and (L3) is highly controversial, but what is perhaps in most doubt is (L4). It is simply not obvious that this premise is true, and Lowe gives no further argument for it. If we are forced to choose between this premise and (2), given there is no further argument for either, it appears that intuition would side with (2), and if there are no further ways to decide between them, such as integration with other theories, it seems that we should opt for (2). This, of course, is no strong argument for (2) over (L4). It is merely an observation that Lowe's argument, if it contains such a premise not supported by further argument or intuitive support, cannot defeat metaphysical nihilism and the intuitions backing it, in particular (2).

So in order to show the subtraction argument to be unsound, it seems that the dialectically most effective strategy would be to accept (2) but claim that (2) does not support a claim as strong as (B), which is needed to entail metaphysical nihilism. One sort of objector who maintains this line argues that (2) is true only if there are at least *two* concrete objects; consequently, (2) does not support (B) and the subtraction argument is unsound.

Perhaps this objector has taken a nonsemantic feature of 'There are some *F*s' to be a deductive implication of this sentence.[38] While it may, in some contexts, be misleading to assert 'There are some *F*s' when there is just one *F*, the assertion is still true, since it is entirely consistent to claim: 'There are some *F*s, perhaps just one'. Consider the following scenario. James and John are stranded on an island. They go about searching for food, and find none. Suddenly, James remembers that he brought along a basket containing just one apple. Exuberantly, James says to John, 'I remember now that I have some apples in my basket'. Has James spoken falsely to John? No, we think not. He may have spoken deceptively buoyed by his excitement, but, surely, he has not spoken falsely since he could equally well have said enthusiastically, 'I remember now that I have some apples in my basket', and continued dolefully, 'but I'm afraid there is just one'. So, in our view, it is a mistake to think that 'There are some concrete objects' is true only if there are at least *two* concrete objects since it is consistent to assert: 'There are some concrete objects, but just one'.

---

[36] "Why Is There Anything at All?," *The Possibility of Metaphysics: Substance Identity and Time*, pp. 252–55, "Metaphysical Nihilism and the Subtraction Argument."

[37] This succinct summary of Lowe's argument is Rodriguez-Pereyra's, and Lowe approves of it. See Rodriguez-Pereyra, "Lowe's Argument Against Nihilism," *Analysis*, LX (October 2000): 335–40, p. 335; and Lowe, "Metaphysical Nihilism and the Subtraction Argument," p. 65.

The objector who thinks that the singleton world is an exception to (2) might have something more sophisticated in mind.[39] It is tempting to read (2) as a strict conditional, which gives the modal operator wide scope and thus makes it appear to the metaphysician's eye to be a claim about what possible worlds there are. And certainly our initial explication of Baldwin's subtraction metaphor made it into such a claim. But then whether (2) applies to the singleton world as well as multiply occupied worlds looks like a matter for modal metaphysics, for our modal metaphysics might well treat the singleton world as a special case. Consequently, the subtraction argument would not give a reason for believing in metaphysical nihilism which is independent of modal metaphysics, as we are trying to argue.

One problem with this line of objection is that it fails to distinguish adequately two quite distinct types of modal theory. A philosopher might give a theory of the metaphysics of unactualized possibility, a theory of what kinds of things possible worlds are. For example, Lewis's famous claim that possible worlds are maximally spatiotemporally interrelated sums of individuals is just such a theory.[40] Alternatively, a philosopher interested in modality might give a theory of what kind of situation is and not possible, a theory of what possible worlds *there are*. For example, Lewis's Principle of Recombination is part of such a theory (*ibid.*, p. 87). While these two theories need to be consistent, neither is prior to or more fundamental than the other. So when they conflict, the resolution of that conflict is just another matter of complex theory choice. (2) is an intuition which supports a claim in the theory of what possibilities *there are*, namely (B), which entails metaphysical nihilism. If some theory of what possible worlds *are* conflicts with metaphysical nihilism, we need to decide between the two theories on the basis of the grounds for each and the costs of rejecting either. It is not obvious that such a decision will always go against (B), and until we have the conflicting theory on the table, we have prima facie grounds for accepting (B).

---

[38] Lepore maintains that those who think that 'Something is a dog' and 'There are dogs' have different truth conditions because the former is true just in case there is at least one dog but the latter is true just in case there are at least two dogs, have taken a conventional implicature of the latter to be a deductive implication. We have adapted Lepore's explanation to respond to those who think that (1) and (2) do not entail (3) because (2) is true only if there are at least two concrete objects. However, we do not think that the rule someone violates when she uses 'some *F*s' while knowing there to be only one *F* is at root conversational. Rather, it is stylistic. If someone knows that there is only one apple, but claims to have some apples, she faces a problem when she wants to say that her apple is not ripe. 'They are not ripe' does imply more than one, but the alternative, 'I have some apples and it is not ripe', is stylistically unacceptable. See Ernest Lepore, *Meaning and Argument* (Malden, MA: Blackwell, 2000), p. 153, note 10.

[39] Paseau has taken the following position in correspondence.

A second problem with this objection is that it makes the modal intuition which supports (2) an intuition about possible worlds. Talk of possible worlds is a theoretical tool for explaining the semantics of modal claims, not a description of our intuitive understanding of modality, which seems in many cases to concern the modal properties of objects. For example, the intuitions that Fido could not have been a cat but he could have been smaller, or that Mary could not have been born a boy but she could have had black hair, are not most naturally construed as being about possible worlds. Rather, it seems best to see them as distinguishing different properties on the basis of their modal character. Of course, they may be equivalent to claims about possible worlds, but that is a matter for philosophical debate.[41] Similarly for the more abstract modal property of being an independent thing, of not depending for its existence upon anything else, the property traditionally captured by the term 'substance'; while possessing such a property may be equivalent to a claim about possible worlds, and the semantics may exploit this equivalence, it does not look plausible to say that this is the most natural way to construe the claim that cabbages and kings are substances but their qualities are not. Whether claims about the modal properties of objects are primi-

---

[40] *On the Plurality of Worlds*, p. 2.

[41] For a challenge to the assumption that claims about essence are equivalent to claims about possible worlds, see Kit Fine, "Essence and Modality," in Tomberlin, ed., pp. 1–16.

CONCLUSION

After sifting through various (mis)interpretations of the subtraction argument, we have given a valid form of it. Additionally, we have identified the modal intuitions that support the argument's premises, and we have argued that we should accept those intuitions. In identifying these intuitions, we have made clear that they concern the modal properties of objects and not possible worlds; from this we have argued that the subtraction argument cannot be deemed to be question begging. We conclude then that the subtraction argument is persuasive.

Now if this conclusion is correct, then whether there might have been nothing is not settled by a pre-existing modal theory. Rather, this question is settled by the subtraction argument. In subsequent work,[43] we show that the leading theories of modality, in particular Lewis's genuine modal realism and Armstrong's combinatorialism, are not directly inconsistent with metaphysical nihilism. This is so because each of these theories can accommodate the subtraction argument and the intuitions which drive it, but not without certain costs elsewhere in metaphysics. So, even if the question of whether metaphysical nihilism is true must be settled by a pre-existing theory of possible worlds, as Armstrong maintains, none of the leading theories settles it. Fortunately, the subtraction argument settles what these theories cannot, namely, that metaphysical nihilism is true.

DAVID EFIRD
TOM STONEHAM

The University of York

---

empty universe is or is not possible is, and should be, spoils for the victor—*A Combinatorial Theory of Possibility*, p. 64.

We say the opposite: a theory of possible worlds should be accepted only if it is consistent with metaphysical nihilism since we think that the subtraction argument is valid and both its premises express common modal intuitions. Theory choice, as Quine has emphasized, often involves trade-offs among simplicity, coherence, and conservation of old doctrine, among other things. One may, of course, take the intuitions backing the subtraction argument less seriously than we do and take the virtues of a theory that render metaphysical nihilism false more seriously than we do.

[43] "Genuine Modal Realism and the Empty World," *European Journal of Analytic Philosophy* (forthcoming), and "Combinatorialism and the Possibility of Nothing" (manuscript).

APPENDIX: THE MECHANICS OF THE SUBTRACTION ARGUMENT

Without assuming that in general the cardinality of a world is a determinate matter, we can say that there are a finite number of concrete objects at a world if, for some natural number n, there are at least 1 and at most n concrete objects. Without loss of generality, let $n=2$:

(A1) $\exists w \exists x \exists y((E!xw \wedge E!yw) \wedge \forall z(E!zw \supset (z = x \vee z = y)))$

Call that world 'A' and the objects 'a' and 'b'.

(2) $E!aA \wedge E!bA \wedge \forall z(E!zA \supset (z = a \vee z = b)$

All concrete objects are necessarily subtractable:

(B) $\forall w_1 \forall x(E!xw_1 \supset \exists w_2(\neg E!xw_2 \wedge \forall y(E!yw_2 \supset E!yw_1)))$

Instantiate the leading quantifiers in (B):

(4) $(E!aA \supset \exists w_2(\neg E!aw_2 \wedge \forall y(E!yw_2 \supset E!yA)))$

From (2) and (4) by modus ponens:

(5) $\exists w_2(\neg E!aw_2 \wedge \forall y(E!yw_2 \supset E!yA))$

Call that world 'B':

(6) $\neg E!aB \wedge \forall y(E!yB \supset E!yA)$

Suppose something 'c' exists at B:

(7) $E!cB$

Instantiate the second conjunct of (6):

(8) $E!cB \supset E!cA$

From (2), (7) and (8) by universal instantiation and two applications of modus ponens:

(9) $c=a \vee c=b$

From (6), (7) and (9) by Leibniz's Law and disjunction elimination:

(10) $c=b$

Discharging (7) by conditional proof and then universal introduction gives:

(11) $\forall x(E!xB \supset x=b)$

Instantiate the leading quantifiers in (B) again:

(12) $(E!bB \supset \exists w_2(\neg E!bw_2 \wedge \forall y(E!yw_2 \supset E!yB)))$

By LEM:

(13) $E!bB \vee \neg E!bB$

Suppose b does exist at B:

(14) $E!bB$

Then from (12) by modus ponens:

(15) $\exists w_2(\neg E!bw_2 \wedge \forall y(E!yw_2 \supset E!yB))$

Call that world 'C'. Then:

(16) $\neg E!bC \wedge \forall y(E!yC \supset E!yB)$

From the second conjunct, (11) and the transitivity of the conditional:

(17) $\forall y(E!yC \supset x{=}b)$

Suppose for reductio that something exists at C:

(18) $\exists x E!xC$

From (17) and (18):

(19) $E!bC$

Which contradicts (16), so we reject (18):

(20) $\forall x \neg E!xC$

And by existential generalization on (20):

(MN) $\exists w \forall x \neg \mathrm{E}!xw$.

Going back to (13), now suppose that b does not exist at B:

(21) $\neg E!bB$

Suppose for reductio that something exists at B:

(22) $\exists x E!xB$

From (11) and (22):

(23) $E!bB$

Which contradicts (21), so we reject (22):

(24) $\forall x \neg E!xB$

And by existential generalization on (24):

(MN) $\exists w \forall x \neg E!xw$.

By simple dilemma on (13):

(MN) $\exists w \forall x \neg E!xw$.

Q.E.D.

NEW BOOKS: TRANSLATIONS

AGAMBEN, GIORGIO: *State of Exception.* KEVIN ATTELL, trans. Chicago: University Press, 2005. vii, 95 p. Paper $12.00.

AVERROES: *Averroes on Plato's Republic.* RALPH LERNER, trans. Ithaca, NY: Cornell, 1974. xxix, 176 p. Paper $16.95.

BARTHES, ROLAND: *The Neutral: Lecture Course at the College de France (1977–1978).* ROSALIND KRAUSS and DENIS HOLLIER, trans. New York: Columbia, 2005. xxix, 280 p. Cloth $29.50.

BATAILLE, GEORGES: *The Cradle of Humanity: Prehistoric Art and Culture.* MICHELLE KENDALL and STUART KENDALL, trans. New York: Zone, 2005. 210 p. Cloth $28.95.

BLANCHOT, MAURICE: *Lautréamont and Sade.* STUART KENDALL and MICHELLE KENDALL, trans. Stanford: University Press, 2004. 190 p. Cloth $50.00, paper $19.95.

BRUNKHORST, HAUKE: *Solidarity: From Civic Friendship to a Global Legal Community.* JEFFREY FLYNN, trans. Cambridge: MIT, 2005. xxv, 262 p. Cloth $42.50.

DERRIDA, JACQUES, and ELISABETH ROUDINESCO: *For What Tomorrow . . .: A Dialogue.* JEFF FORT, trans. Stanford: University Press, 2004. x, 241 p. Cloth $49.50, paper $19.95.

ERASMUS: *Collected Works of Erasmus: Adages III iv 1 to IV ii 100.* DENIS L. DRYSDALL, trans. Toronto: University Press, 2005. xii, 592 p. Cloth $150.00.

FERRY, LUC: *What Is the Good Life?* LYDIA G. COCHRANE, trans. Chicago: University Press, 2005. x, 320 p. Cloth $45.00.

GARFF, JOAKIM: *Søren Kierkegaard: A Biography.* BRUCE H. KIRMMSE, trans. Princeton: University Press, 2005. xxv, 867 p. Cloth $35.00.

GOMBROWICZ, WITOLD: *A Guide to Philosophy in Six Hours and Fifteen Minutes.* BENJAMIN IVRY, trans. New Haven: Yale, 2004. 111 p. Cloth $22.00; *Polish Memories.* BILL JOHNSTON, trans. New Haven: Yale, 2004. 191 p. Cloth $22.00.

HEGEL, G.W.F.: *Hegel's Philosophy of Nature.* A.V. MILLER, trans. New York: Oxford, 2004. xxxi, 450 p. Paper $35.00; *Hegel's Preface to the Phenomenology of Spirit.* YIRMIYAHU YOVEL, trans. Princeton: University Press, 2005. xii, 223 p. Cloth $19.95.

HEIDEGGER, MARTIN: *Introduction to Phenomenological Research.* DANIEL O. DAHLSTROM, trans. Bloomington: Indiana UP, 2005. xiv, 252 p. Cloth $39.95.

HÖSLE, VITTORIO: *Morals and Politics.* STEVEN RENDALL, trans. Notre Dame: University Press, 2004. xxi, 991 p. Cloth $60.00.

KANT, IMMANUEL: *Groundwork For the Metaphysics of Morals.* THOMAS K. ABBOTT, trans. LARA DENIS, ed. Orchard Park, NY: Broadview, 2005. 209 p. Cloth $40.00, paper $7.95; *Notes and Fragments.* CURTIS BOWMAN, PAUL GUYER, and FREDERICK RAUSCHER, trans. New York: Cambridge, 2005. xxx, 663 p. Cloth $140.00.

LOSURDO, DOMENICO: *Hegel and the Freedom of Moderns.* MARELLA MORRIS and JON MORRIS, trans. Durham, NC: Duke, 2004. xix, 379 p. Cloth $89.95, paper $24.95.

NEW IN PAPERBACK **PENGUIN GROUP (USA)**

A TERRIBLE LOVE OF WAR
JAMES HILLMAN
A profound examination of the roots of humanity's unavoidable obsession with war. "Important reading for our time, as we try to make sense of our errors."—*San Francisco Chronicle.*
Penguin 272 pp. 0-14-303492-8 $15.00

MULTITUDE
War and Democracy in the Age of Empire
MICHAEL HARDT and ANTONIO NEGRI
"Hardt and Negri's impressive work sheds politics for philosophy and factionalism for foresight. A rare and exciting work of synthesis...blends some of the most cutting-edge scholarly work on globalization into a relatively accessible package."—*Booklist.*
Penguin 448 pp. 0-14-303559-2 $16.00
Available August 2005

CATCH AND RELEASE
Trout Fishing and the Meaning of Life
MARK KINGWELL
A philosopher-angler's funny and provocative meditation on his twin passions. "Illustrates deep thoughts on truth, beauty and thinking itself with laugh-out-loud yarns of male bonding."—*The Boston Globe.*
Penguin 256 pp. 0-14-303514-2 $15.00

NEW FROM PENGUIN CLASSICS

THE LAWS / PLATO
Translated with an Introduction and Notes by Trevor J. Saunders
New Preface by Richard Stalley
Penguin Classics 560 pp. 0-14-044984-1 $14.00 *July 2005*

MENCIUS
Translated with an Introduction and Notes by D. C. Lau
Lau's lucid translation has been updated.
Penguin Classics 304 pp. 0-14-044971-X $15.00 *July 2005*

ETHICS / BENEDICT DE SPINOZA
Translated by Edwin Curley
Introduction by Stuart Hampshire
For the first time in Penguin Classics, the standard translation of Spinoza's greatest work.
Penguin Classics 208 pp. 0-14-043571-9 $14.00 *August 2005*

PENGUIN GROUP (USA)
Academic Marketing Dept., 375 Hudson St., NY, NY 10014 www.penguin.com/academic

Graduate Faculty Philosophy Journal

The **Graduate Faculty Philosophy Journal**, published biannualy in association with the Department of Philosophy, New School for Social Research, is a forum for historical and contemporary issues in philosophy. The *Journal* also publishes an annual special issue dedicated to investigating the historical development of concepts and problems in the history of philosophy and their ongoing relevance to issues in contemporary philosophy.

Vol. 26, No. 1

Philosophy of Dialogue

Hannah Arendt/Reiner Schürmann Memorial Symposium

Charles Kahn	The Philosophical Significance of the Dialogue Form for Plato
Vittorio Hösle	Self-transcendence of Philosophical Dialogues in Plato and Cicero
Caryl Emerson	The Dialogic Word in Literary Art: What Sort of Fiction is This?
Dmitri Nikulin	Dialogue vs. Discourse: On the Possibility of Disagreement in Human Communication
Rainer Forst	Between Discourse and Dialogue: The Normativiy of Morality
Georgia Warnke	Interpretive Democracy

Additional Contributions

Jean-Luc Marion	Phenomenon and Event
Thomas Tam	The Death of Art: Bataille, Malraux, Hegel
Drucilla Cornell	Who Bears the Right to Die

Reflections on Jacques Derrida

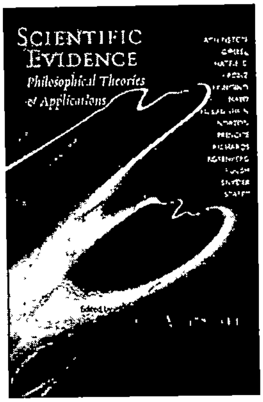
by **Richard J. Bernstein** and **Agnes Heller**

All communications should be addressed to the Editor, **Graduate Faculty Philosophy Journal**, Department of Philosophy, New School for Social Research, 65 Fifth Avenue, New York, NY 10003. The *Journal* is biannual. Domestic rates: Individuals: $25.00/year; Students: $15.00/year; Institutions: $50.00/year.

SCIENTIFIC EVIDENCE

Philosophical Theories and Applications

EDITED BY PETER ACHINSTEIN



ACHINSTEIN
GIMBEL
HATFIELD
KRONZ
LONGINO
MAYO
MCLAUGHLIN
NORTON
PRINCIPE
RICHARDS
ROSENBERG
ROUSH
SNYDER
STALEY

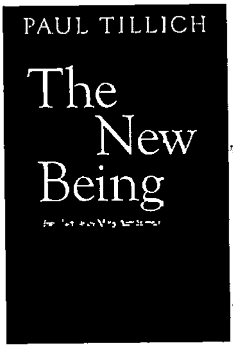
"Leading philosopher of science Peter Achinstein has put together a first-rate collection on the problem of scientific evidence."
—MICHAEL RUSE,
Florida State University

Peter Achinstein has gathered prominent philosophers and historians of science for critical and lively discussions of both general questions about the meaning of evidence and specific ones about evidence for particular scientific theories.
$49.95 hardcover

THE JOHNS HOPKINS UNIVERSITY PRESS
1-800-537-5487 • www.press.jhu.edu

The New Being
BY PAUL TILLICH
Introduction by Mary Ann Stenger
$12.95 paper

Jonathan Edwards
BY PERRY MILLER
Introduction by John F. Wilson
$16.95 paper



UNIVERSITY
OF NEBRASKA
PRESS
www.nebraskapress.unl.edu

You can now subscribe to the

THE JOURNAL OF PHILOSOPHY

using

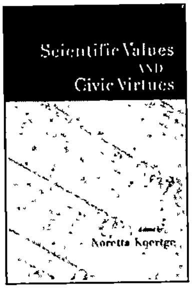
PayPal

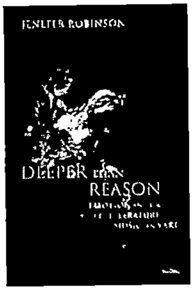
New Rates January 2005

Institutional	$100
Individual	$45
Student/Retired	$20

To learn more about this easy new option, click on the PayPal icon at www.jphil.org

New from
Oxford





Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. Visit our website at www.oup.com/us

OXFORD
UNIVERSITY PRESS

New in Paperback!
The Heirs of Plato
A Study of the Old Academy (347–274 BC)
JOHN DILLON
"To write a book on the Old Academy is a brave undertaking...It is a substantial contribution, and has certainly helped me clarify my own thinking on many of these issues."—*Ancient Philosophy*
paper June 2005 $35 00

Spinoza and Spinozism
An Introduction
STUART HAMPSHIRE
Hampshire's classic 1951 book *Spinoza* remains the best introduction to this thinker, and it is reprinted here. But what gives particular interest to this new volume is the first publication of Hampshire's last work "Spinoza and Spinozism," an extended presentation of a Spinozist philosophical worldview. Hampshire's influential 1962 essay "Spinoza and the Idea of Freedom" is also included.
June 2005 paper $24.95 cloth $74.00

Scientific Values and Civic Virtues
Edited by NORETTA KOERTGE
This volume of contributed essays takes an affirming, positive view of the relationship between the values embodied in science (such as honesty and precision of language) and the nature of civil society. The contributors—who include philosophers, political scientists, feminist theorists, physicists, and engineers—argue that science can broadly inspire the civic virtues of an educated and tolerant global enterprise dedicated to the common good.
June 2005 paper $25.00 cloth $70.00

Deeper than Reason
Emotion and Its Role in Literature, Music, and Art
JENEFER ROBINSON
Deeper than Reason takes the insights of modern scientific research on the emotions and uses them to illuminate questions about our emotional involvement with the arts. Written in a clear and engaging style, Robinson's book will make fascinating reading for anyone interested in the emotions and how they work, as well as anyone engaged with the arts and aesthetics, especially with questions about emotional expression in the arts, emotional experience of art forms, and more generally, artistic interpretation.
June 2005 $45.00

Faith and Reason
Second Edition
RICHARD SWINBURNE
Richard Swinburne presents a new edition of one of his classic works on philosophical theology. While maintaining the same structure and conclusions as the original, this second edition has been substantially rewritten, both in order to relate its ideas more closely to those of classical theologians and philosophers and to respond to more recent views.
June 2005 paper $34.95 cloth $99.00 JKC.040505

CAMBRIDGE

OUTSTANDING SCHOLARSHIP

The Cambridge Companion to Arabic Philosophy
Edited by Peter Adamson and Richard C. Taylor
Cambridge Companions to Philosophy
$70.00: Hardback: 0-521-81743-9: 466pp
$29.99: Paperback: 0-521-52069-X

Is There a Right of Freedom of Expression?
Larry Alexander
Cambridge Studies in Philosophy and Law
$65.00: Hardback: 0-521-82293-9: c.216pp
$23.00: Paperback: 0-521-52984-0

Aquinas: Disputed Questions on the Virtues
Edited by E.M. Atkins and Thomas Williams
Cambridge Texts in the History of Philosophy
$75.00: Hardback: 0-521-77225-7: 343pp
$26.99: Paperback: 0-521-77661-9

Corporate Integrity
Rethinking Organizational Ethics and Leadership
Marvin T. Brown
$75.00: Hardback: 0-521-84481-9: 284pp
$34.99: Paperback: 0-521-60657-8

The Cambridge Companion to Anselm
Edited by Brian Davies and Brian Leftow
Cambridge Companions to Philosophy
$65.00: Hardback: 0-521-80746-8: 248pp
$29.99: Paperback: 0-521-00205-2

The Selfish Meme
A Critical Reassessment
Kate Distin
$60.00: Hardback 0-521-84452-5: 238pp
$21.99: Paperback: 0-521-60627-6

Now in paperback...

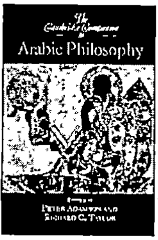
Lack of Character
Personality and Moral Behavior
John M. Doris
$26.99: Paperback: 0-521-60890-2: 284pp

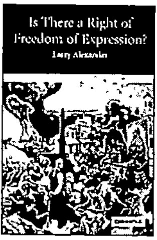
Newton: *Philosophical Writings*
Edited by Andrew Janiak
Cambridge Texts in the History of Philosophy
$60.00: Hardback: 0-521-83122-9: 190pp
$23.99: Paperback: 0-521-53848-3

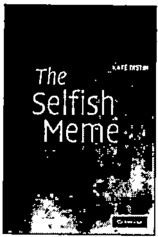
Medieval Islamic Philosophical Writings
Edited by Muhammad Ali Khalidi
Cambridge Texts in the History of Philosophy
$65.00: Hardback: 0-521-82243-2: 236pp
$25.99: Paperback: 0-521-52963-8

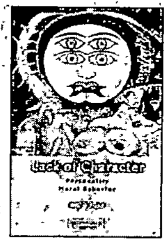
Perfection and Disharmony in the Thought of Jean-Jacques Rousseau
Jonathan Marks
$60.00: Hardback: 0-521-85069-X: 200pp

The Cambridge Companion to Chomsky
Edited by James McGilvray
$75.00: Hardback: 0-521-78013-6: 346pp
$29.99: Paperback: 0-521-78431-X


The Cambridge Companion to Arabic Philosophy
Peter Adamson and Richard C. Taylor







www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

FROM CAMBRIDGE

Now in paperback...

Faking It
William Ian Miller
$18.99: Paperback: 0-521-61370-1: 302pp

Lessing: *Philosophical and Theological Writings*
Edited by H.B. Nisbet
Cambridge Texts in the History of Philosophy
$70.00: Hardback: 0-521-83120-2: 280pp
$27.99: Paperback: 0-521-53847-5

The Persistence of Subjectivity
On the Kantian Aftermath
Robert Pippin
$75.00: Hardback: 0-521-84858-X: 368pp
$28.99: Paperback: 0-521-61304-3

How the Cold War Transformed Philosophy of Science
To the Icy Slopes of Logic
George A. Reisch
$70.00: Hardback: 0-521-83797-9: 432pp
$26.99: Paperback: 0-521-54689-3

Wittgenstein's *Philosophical Investigations*
An Introduction
David G. Stern
Cambridge Introductions to Key Philosophical Texts
$65.00: Hardback: 0-521-81442-1: 224pp
$23.99: Paperback: 0-521-89132-9

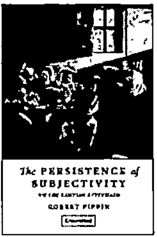
Evidence and Faith
Philosophy and Religion since the Seventeenth Century
Charles Taliaferro
The Evolution of Modern Philosophy
$75.00: Hardback: 0-521-79027-1: 472pp
$29.99: Paperback: 0-521-79375-0

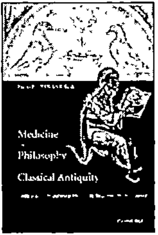
Medicine and Philosophy in Classical Antiquity
Doctors and Philosophers on Nature, Soul, Health and Disease
Philip J. van der Eijk
$95.00: Hardback: 0-521-81800-1: 418pp

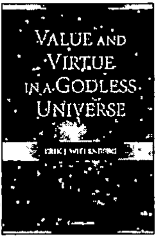
Quality of Life and Human Difference
Genetic Testing, Health Care, and Disability
Edited by David Wasserman, Jerome Bickenbach, and Robert Wachbroit
Cambridge Studies in Philosophy and Public Policy
$70.00: Hardback: 0-521-83201-2: 288pp
$24.99: Paperback. 0-521-53971-4

Value and Virtue in a Godless Universe
Erik J. Wielenberg
$60.00: Hardback: 0-521-84565-3: 204pp
$20.99. Paperback: 0-521-60784-1

Prices subject to change.







www.cambridge.org/us/philosophy/

CAMBRIDGE
UNIVERSITY PRESS

The Roots of Evil

John Kekes

"Evil is here defined as harm in excess and examined with the lucidity and clarity that distinguish John Kekes's books. His focus is on the thrill of evil, and his examples are stunning. This is a work of philosophy for every serious reader."

—Harvey C. Mansfield, Harvard University

$29.95 cloth

The Art of Life

John Kekes

*"*The Art of Life *demonstrates unequivocally the value of rigorous philosophical reflection on something that we all want, spend most of our time and energy trying to achieve, fret about endlessly, but understand only dimly, namely, a good life."*

—*The Journal of Ethics*

$19.95 paper

Pythagoras

His Life, Teaching, and Influence

Christoph Riedweg, translated from the German by Steven Rendall

"Riedweg gives a rich picture of Pythagoras as a figure of legend, with abundant details from the Neoplatonist 'Lives' of Pythagoras, including many supernatural exploits and edifying stories."

—Charles H. Kahn, author of *Pythagoras and the Pythagoreans*

$29.95 cloth

The Word Made Self

Russian Writings on Language, 1860–1930

Thomas Seifrid

"This is the first comprehensive examination in English of the rich and multifarious body of discourses on language produced in Russia over a period of some seventy years, from the abolition of serfdom to the rise of Stalinism."

—Galin Tihanov, Lancaster University

$45.00 cloth

The Philosophy of the Commentators, 200–600AD

A Sourcebook

Richard Sorabji

Volume 1: Psychology (with Ethics and Religion)	*$39.50 paper*
Volume 2: Physics	*$39.50 paper*
Volume 3: Logic and Metaphysics	*$39.50 paper*

Interpretive Reasoning

Laurent Stern

"This deft and adventurous book about the nature and problems of interpretation displays a rare combination of cultural breadth and analytical acuity."

—Garry L. Hagberg, Bard College

$39.95 cloth

Virtue Ethics, Old and New

edited by Stephen M. Gardiner

*"*Virtue Ethics, Old and New *makes significant original contributions and offers useful continuations of current topics."*

—Michael Stocker, Syracuse University

$49.95 cloth, $19.95 paper

CORNELL UNIVERSITY PRESS

www.cornellpress.cornell.edu | 1-800-666-2211